Modeling and matching digital data marketplace policies

Shakeri, S.; Maccatrozzo, V.; Veen, L.; Bakhshi, R.; Gommans, L.; de Laat, C.; Grosso, P.

DOI
10.1109/eScience.2019.00078

Publication date
2019

Document Version
Final published version

Published in
IEEE 15th International Conference on eScience

Modeling and Matching Digital Data Marketplace Policies

Sara Shakeri
Systems and Networking Lab
University of Amsterdam
Amsterdam, The Netherlands
s.shakeri@uva.nl

Valentina Maccatrozzo
Netherlands eScience Center
Amsterdam, The Netherlands
v.maccatrozzo@esciencecenter.nl

Lourens Veen
Netherlands eScience Center
Amsterdam, The Netherlands
l.veen@esciencecenter.nl

Rena Bakhshi
Netherlands eScience Center
Amsterdam, The Netherlands
r.bakhshi@esciencecenter.nl

Leon Gommans
Systems and Networking Lab
University of Amsterdam
Amsterdam, The Netherlands
l.gommans@uva.nl

Cees de Laat
Systems and Networking Lab
University of Amsterdam
Amsterdam, The Netherlands
delaat@uva.nl

Paola Grosso
Systems and Networking Lab
University of Amsterdam
Amsterdam, The Netherlands
p.grosso@uva.nl

***Abstract*—Recently, Digital Data Marketplaces (DDMs) are gaining wide attention as a sharing platform among different organizations. That is due to the fact that sharing the information and participating in research collaborations play an important role in addressing multiple scientific challenges. To increase trust among participating organizations multiple contracts and agreements should be established in order to determine regulations and policies about who has access to what. Describing these agreements in a general model to be applicable in different DDMs is of utmost importance. In this paper, we present a semantic model for describing the access policies by means of semantic web technologies. In particular, we use and extend the Open Digital Rights Language (ODRL) to describe the pre-established agreements in a DDM.**

***Index Terms*—Digital Data Marketplaces, Sharing policies, Semantic modeling, ODRL, Archetype**

I. Introduction

In recent years, digital collaboration and data sharing among different scientific and commercial organizations have become a substantially interesting concept. There are a lot of organizations willing to share their data with other parties in sharing platforms to address different scientific challenges by analyzing the shared data [1], [2]. These platforms are called digital data marketplaces (DDMs). In fact, sharing the data with different parties offers long-term benefits for the cooperating organizations [2]. This cooperation among organizations can occur in diverse domains, such as medical, financial, and scientific services. In most cases, the shared data is used for computational and mining processes to take advantage of the generated results. Therefore, the participating organizations in DDMs are willing to use both shared information and shared computational applications. For example, in a medical system in which the healthcare providers want to find the relationship between patient and disease characteristics, the best way is using a shared computational algorithm to analyze the patients' data and medical records and then finding the relationship utilizing the results.

Digital data marketplaces provide a distributed alternative to data silos run by a small number of large organizations. In a distributed system, control over data is kept with the owner or subject of the data, which benefits privacy and reduces market-distorting monopolies. Accordingly, providing a data sharing platform among different participating parties is of paramount importance. However, there are still some challenges about the methods of bringing required trust and security in DDMs as sharing environments [3], [4]. This denotes the importance of arranging appropriate sharing policies in the platform so that digital collaboration becomes possible. By describing and then implementing these policies, the parties can be sure that only the authorized organizations can access their data and this will increase trust in DDMs.

Sharing policies are based on pre-established agreements between participating organizations, for example a GDPR (General Data Protection Regulation) data processing agreement. They determine permission and prohibition access rules related to a specific object in a specific location. The first major step for bringing a secure data sharing platform is describing these rules and regulations in an efficient way. A full description will make the rules more clear and then implementing them in the infrastructure would be more straight-forward.

In the SecConNet project[1], we want to build a data sharing platform to support automatic handling users' requests. We do this by means of Semantic Web technologies. In particular, we use and extend the Open Digital Rights Language (ODRL) [5] to describe the pre-established agreements between participating organizations and to model users' requests.

In this paper, we present:

1) A generic semantic model that leverages the ODRL ontology and extends it specifically for data sharing

This work is supported by the Netherlands eScience Center and NWO under the project SecConNet.

[1]https://www.esciencecenter.nl/project/secconnet

978-1-7281-2451-3/19/$31.00 ©2019 IEEE
DOI 10.1109/eScience.2019.00078

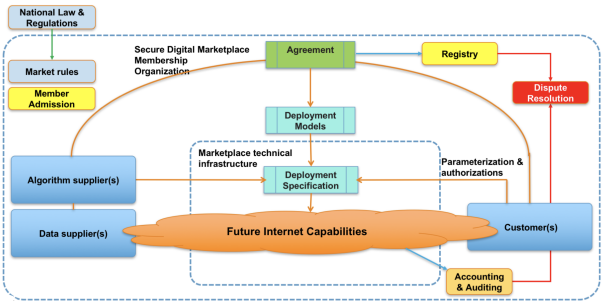

Fig. 1. Secure Digital Data Marketplace framework [6]

applications in DDMs.

2) The deployment of the semantic model to automatic handling of users' requests in the data sharing platform.

The remainder of this paper is organized as follows. In the next section, we present background about the methods of setting up secure data sharing platforms. DDMs architecture will be explained in section III. Then the semantic model and ODRL ontology will be described in section IV, following by representing the matching module in section V. Finally, the paper will be concluded in section VIII.

II. Digital Data Marketplace Platform

With the development of digital data sharing, providing an efficient and secure data sharing platform is a fundamental requirement. There have been intensive efforts for setting up a high-performance infrastructure so that sharing the data among different parties becomes possible. Fig. 1, shows the general architecture of data sharing platforms. The marketplace infrastructure is built based on the agreement which is held between algorithm suppliers, data suppliers, and customers. The agreement can be presented in different deployment models and then it has to be implemented in the marketplace infrastructure. In addition, several other factors like authorization, auditing, and accounting should be taken into consideration. In the remainder of this section, we summarize related projects which aimed to build data sharing platforms for scientific and commercial collaborations. All of these projects are launched in the Netherlands:

Data Logistics for Logistics Data (DL4LD) [6] is going to provide an effective solution that allows organizations to agree on how data is shared and exchanged along with deploying a controllable, enforceable, and goal-oriented method. The project is based on the architecture shown in Fig. 1. It shows different components including agreement, deployment models, suppliers of digital resources, and customers. As depicted in Fig. 1, the agreement must be deployed and mapped on the infrastructure before the exchanging of the digital resources can take place.

Enabling Personalized Interventions (EPI) [7] aims to develop a healthcare platform based upon a secure and trustworthy distributed data infrastructure to create new, actionable, and personalized insights for providers and patients. It benefits from the collaboration of medical professionals, data scientists, ICT-infrastructure experts, and machine learning researchers.

Green Village Data Sharing Platform [8] is a project defined by the Green Village with the cooperation of the ICT innovation department of TUD and SURFsara [9]. It is aimed at providing a reliable and highly available sharing platform to exchange data among different organizations.

The Neutral Logistics Information Platform (NLIP) [10] is a part of the Netherlands' Logistics Top Sector program, a leading platform promoting data exchange in the transport and logistics sector. One of the projects defined in this context is iSHARE [11]. It is an appointment system for identification, authentication, and authorization to share the logistics data in a safe and controlled fashion. This system can be used by all parties which have activity in the logistics sector. Overall, the NLIP project is trying to facilitate the development of tools and digital standards for accessing and sharing data sources in order to eliminate the barriers of data sharing and reduce the pressure on physical infrastructure.

The Digital Object Architecture (DOA) can be considered an early version of a digital data marketplace [12]. The DOA market model is less tightly organized than the DDM: it assumes that individual organizations will offer data objects. These may include documents as well as software, which are registered with a central metadata repository and given a unique identifier. Digital Data Marketplaces include this functionality (possibly implemented in a distributed fashion using a blockchain), but offer more powerful models of collaboration between organizations and data processing. While DOA does assume that operations on data may be performed by a data repository, it does not define how this is done, which restrictions apply, or how individual operations are combined into a distributed workflow.

Virtual organizations are collaborations between organizations in the context of the grid [13]. A virtual organization has access to compute resources, data and software, and makes those available to its members, and is thus similar to a DDM. However, virtual organizations are complex to manage, and proper isolation of users and organizations is difficult to achieve on typical grid infrastructure [14]. The latter issue is now routinely solved using virtualization technologies, and we believe that the use of semantic technologies will help deal with the complexity inherent in large-scale collaborations in limited-trust environments.

Although all the mentioned works explore different methods of providing the sharing platform, in this paper, we specifically concentrate on increasing security and trust in the network by introducing a semantic model to describe the policies in the sharing platform. This substantially makes the collaboration of different parties more efficient.

III. Digital Data Marketplace System Architecture

Digital data sharing and digital collaboration in a DDM rely on the sharing policies and agreements between parties.

Ensuring the implementation of these agreements and policies introduces the need for representing an efficient and secure data sharing system in a DDM. The DDM sharing system architecture offered in this paper is shown in Fig. 2. By construction, it relies on four components: 1) Application; 2) Archetypes; 3) Matching Module; and, 4) Container-based Infrastructure. In the following, we describe each of these components.

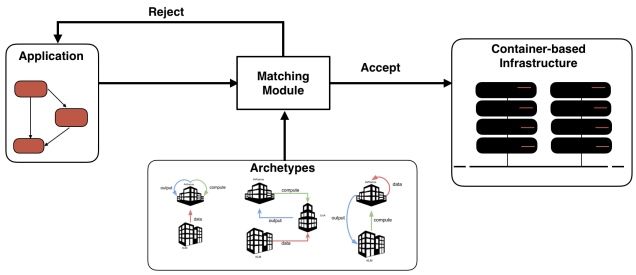

Fig. 2. Digital Data Marketplace system architecture

- **Application** Two kinds of digital resources can be shared in the proposed DDM system.
  - Digital Algorithm: A program that operates on data.
  - Digital Data, including:
    - Input Data: Input data of the algorithm.
    - Output Data: Result of executing the algorithm.

  We consider an application as a set of computational processes in which the operation that the algorithm performs on the input data and the generation of the output is defined. Fig. 3 shows an application encompassing two different computational processes. Here the output of computational process 1 is used as an input for computational process 2.

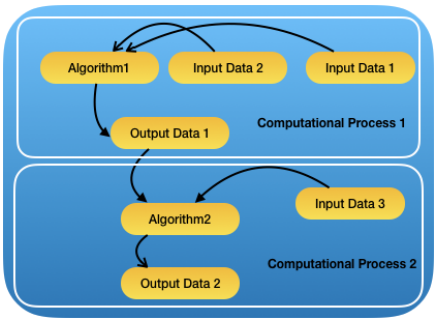

Fig. 3. Digital Data Marketplace application

- **Archetype** A scenario that determines the permitted transmissions of the shared digital resources. Fig. 4 depicts a simple archetype. It shows the transmission of the input data from *KLM* organization to location *trusted*, and the transmission of the algorithm from *Airfrance* organization to location *trusted* is allowed. Also, the output is allowed to be transferred to *Airfrance*. Any other transformation of data and algorithm is denied.

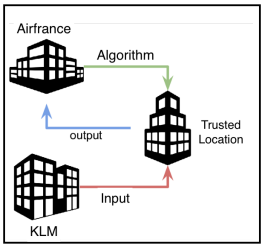

Fig. 4. Digital Data Marketplace archetype

- **Matching Module** As shown in Fig. 2, the matching module automatically verifies whether the requested application is matched with one of the available archetypes in the DDM system or not. Let's consider a scenario in which a user from a specific organization has a request based on application in Fig. 3. If this request matches with one of the available archetypes in the contract, the system will set up the infrastructure based on the archetype and starts performing the request. Otherwise, the request will be rejected.
- **Container-based Network Infrastructure** The platform that provides the connection among different parties. Several infrastructure architectures are possible. We expect that infrastructures based on the use of Linux containers will prove the most flexible. Network implementation of the infrastructure is out of the scope of this paper.

IV. Semantic Model

In this section, we describe the guiding principles of the semantic model, the Open Digital Rights Language (ODRL) ontology and the extensions needed.

The main goal of our model is to describe how digital resources can be transferred and/or shared among different parties in a DDM.

A. Model requirement

In DDMs, contracts are defined between parties in order to regulate the flow of resources (e.g., algorithms, data) among their infrastructures. For instance, party A might want to share its data in the DDM but does not want it to leave its premises. If party B wants to run its algorithm on A's data, it can send its algorithm to A, and, after execution, A will send the results back to B. Of course, A has to trust that B's algorithm is not actually copying the data and B has to trust that A will not copy its algorithm and will return the proper results. In this paper, we assume that parties taking part in a DDM do trust each other and that proper measures have been put in place to make sure trust is not misplaced.

We need a simple language to describe how resources can be shared and used by different parties. ODRL is a W3C recommendation language designed to model permissions, obligations, and prohibitions regarding digital resources. Through these rules, the model allows describing the terms of use and reuse of digital content. This model shares many similarities with our requirements, and we needed only a few

extensions to customize it for our needs. In the next subsection we introduce the ODRL information model[2] and, later, we illustrate the extensions we propose.

B. ODRL Information Model

The ODRL information model allows for a flexible description of policies, by modeling what is allowed and what is not, as well as other terms, requirements, and parties involved. The classes from the ODRL information model that we use are:

- *Asset*: a digital resource, e.g., data, and algorithms.
  - *AssetCollection* (sub-class of *Asset*): a group of Assets with common characteristics. This is used to describe categories of assets. Note that ODRL allows both extensional and intensional definitions of AssetCollections.
- *Action*: an activity performed on an Asset. This is used to describe how the Assets could be shared in the DDM.
- *Rule*: description of an action to be performed over an *Asset*.
  - *Permission* (sub-class of *Rule*): a description about what is allowed to do with an *Asset*.
  - *Duty* (sub-class of *Rule*): a description of an action a party is obliged to perform with an *Asset*.
- *Party*: entities which assume roles in a rule. A party is a member of the DDM, which approved the Agreement.
  - *PartyCollection* (sub-class of *Party*): a group of parties with common characteristics (e.g. employees of the same company).
- *Constraint*: Refinement of an Action, a Rule, and a Party/Asset collection. This is used, for instance, to refine the movement of an asset to a specific digital location.
- *Policy*: a group of rules. Policies are defined by means of rules about the usage of digital resources (i.e., assets).
  - *Agreement* (sub-class of *Policy*): granting of Rules from assigner to assignee parties. This class is used to describe the contract signed by the parties in the DDM.

C. SecConNet semantic model

In Fig. 8, we present an example of a policy, representing the archetype described in Fig. 4. The figure is divided into three main boxes, from top to bottom: Algorithm, Input, and Output. In every box, we can see the Asset description on the left (the Asset Collection box). Asset Collections are used to group together assets which share the same rule set. In particular, given the fact that contracts in DDMs define how assets can be used by the parties and that there could possibly be an unlimited number of parties, with an unlimited number of assets to share, it would be rather inconvenient to define rules for every asset in the DDM. To overcome this issue, we propose to define categories of assets, i.e. Asset Collections, hence a party needs only to declare which category its assets belong to. Every box includes also the description of the actions allowed on that specific asset. Most of these actions are described by a rule of the type permission. In the Output box, there is a rule of the type duty. This rule describes the fact that the output generated by the algorithm has to be moved to location *AirFrance*.

Data in DDMs is shared for use as input to other parties' algorithms. While the concept of output (defined as the asset which is created from the output of an action, see Output box in Fig. 8) is included in ODRL, the concept of input is missing. When the target of action is an algorithm and the action is "execute", we have no way to define the data used as input. So, we extend the ODRL model by adding the "input" property, as we show in Fig. 8 in the input box.

In the SecConNet semantic model, we also use other models, such as the PROV Ontology (PROV-O)[3] to describe and record every action performed in the DDM, FOAF[4] to describe parties, and the Data Catalog Vocabulary (DCAT)[5] to describe datasets. However, these models are not included in Fig. 8, as the focus of this paper is on the modeling of the policy.

V. Matching module

The matching module's main goal is to allow for automatic management of user requests. In the DDM, users can submit requests to use specific datasets or algorithms, specifying the location of execution. When sending an application, users, have to specify:

- the dataset they want to use;
- the algorithm they want to use;
- the location of execution of the application;
- the location where the results of the application have to be sent.

Finally, using SPARQL queries, the matching module will verify whether the request is doable and approve or reject it.

While the main goal of the matching model is to verify the applicability of the request, it can be easily extended and used to guide the user in the submission of the request. For instance, considering the user's credentials, only the datasets that the user is allowed to use will be listed. Once the dataset is selected, only allowed locations will be shown, and so on, until all the fields will be filled in.

We illustrate the behavior of the matching module with the following example.

A. Example

We show the functioning of the matching module by testing the applicability of the archetype described in Fig. 4. As a contract defining the DDM, we use the one described in Fig. 8. The application request consists of:

- use a *KLM* dataset (Data1);
- use an *Airfrance* algorithm (Algorithm1);
- *trusted* is the location of execution;

[2]https://www.w3.org/TR/odrl-model/

[3]https://www.w3.org/TR/prov-o/

[4]http://xmlns.com/foaf/spec/

[5]https://www.w3.org/TR/vocab-dcat/

```
SELECT ?location
      WHERE {SECCONET:algorithmAsset/Algorithm1 PROV:atLocation ?location.
      }

SELECT ?location
      WHERE {SECCONET:dataAsset/Data1 PROV:atLocation ?location.
      }
```

Fig. 5. SPARQL queries to verify where Algorithm1 and Data1 are located.

- the results need to be sent to *Airfrance*.

With the help of Alg 1, we guide the reader through the steps the module performs. The first step is to verify where the dataset and the algorithm to be used in the application are located (see Alg 1 line 8). In Fig. 5, we show the SPARQL queries to find the location of Algorithm1 and Data1. Location of all the assets in the DDM is recorded using the property *PROV:atLocation*. In case the assets are not in the requested execution location, we need to verify whether they could be moved (see Alg 1 line 3). In Fig. 6, we show the SPARQL queries to verify whether the assets can be moved to the requested location of execution, after being copied.

Alg 1: Matching Algorithm

Input: Algorithm1, Data1, Execution Location, Output Location

1 **Function** *moveAllowed(Asset, Location)*
2 **if** *location(Asset)* $\neq$ *Location* **then**
3 **return** *whether Asset may be moved to Location*
4 **end**
5 **return** (True)
6 **end**
7 Output1 = outputOf(Algorithm1, Data1);
8 **if** *moveAllowed(Algorithm1, Execution Location) and moveAllowed(Data1, Execution Location) and moveAllowed(Output1, Output Location)* **then**
9 Accept the Request;
10 **else**
11 Reject the Request;
12 **end**

The last step for completing the matching is to verify whether the results could be moved to the location *Airfrance* (see Alg 1 line 7-8). In Fig. 7 we show the SPARQL query to verify whether Algorithm1 can be executed in location *trusted* with input Data1 and if the results could be moved to location *Airfrance*.

VI. Discussion

The example presented in section V-A shows how we use the semantic model of a policy to perform an automatic matching of the application request with the rules of the DDM.

Mello et al. [15] provides three requirements for infrastructures for sharing clinical trial data: 1) the system must provide sufficiently broad access, 2) it must ensure accountability of all parties involved, 3) and it must be practicable. As both commercial and privacy aspects play a role for clinical trial

```
SELECT ?moveAlgorithm
         WHERE {
          ?ruleCopy ODRL:action "reproduce".
          ?ruleCopy ODRL:target SECCONET:algorithmAsset/Algorithm1.
          ?ruleCopy ODRL:output ?moveAlgorithm.
          ?rule ODRL:target ?moveAlgorithm.
          ?rule ODRL:action "move".
          ?rule ODRL:constraint ?constraint.
          ?constraint ODRL:virtualLocation SECCONET:location/trusted.
         }

SELECT ?moveData
         WHERE {
          ?ruleCopy ODRL:action "reproduce".
          ?ruleCopy ODRL:target SECCONET:dataAsset/Data1.
          ?ruleCopy ODRL:output ?moveData.
          ?rule ODRL:target ?moveData.
          ?rule ODRL:action "move".
          ?rule ODRL:constraint ?constraint.
          ?constraint ODRL:virtualLocation SECCONET:location/trusted.
         }
```

Fig. 6. SPARQL queries to verify whether it is possible to move Algorithm1 and Data1 to the requested location.

```
SELECT ?ruleMoveOutput
         WHERE {
          ?rule ODRL:action "execute".
          ?rule ODRL:target SECCONET:AlgorithmAsset/Algorithm1.
          ?rule ODRL:output ?output.
          ?rule ODRL:input SECCONET:DataAsset/Data1.
          ?rule ODRL:constraint ?constraint .
          ?constraint ODRL:virtualLocation SECCONET:location/trusted.
          ?ruleMoveOutput ODRL:action "move".
          ?ruleMoveOutput ODRL:target ?output.
          ?ruleMoveOutput ODRL:constraint ?constraint2.
          ?constraint2 ODRL:virtualLocation SECCONET:location/airfrance.
         }
```

Fig. 7. SPARQL query to verify whether it is possible to move the results of the application to the requested location.

data, we believe that their criteria apply more broadly, and provide a good reference to measure our approach against.

Access to the DDM is in practice mostly a matter of trust and economic concerns, and less of a technical issue. However, a system that can accommodate complicated requirements with respect to access to and use of data and algorithms will arguably allow more parties to participate. ODRL is a powerful right description language, and the use of semantic technology makes it easy to extend the ontology if needed. It is flexible enough to be extended to support different types of policies when it is necessary.

From a technical perspective, a DDM consists of users, data and algorithm providers, and infrastructure providers. To ensure accountability of users, requests need to be matched against the archetypes specified in the contracts as demonstrated above. Some improvement can still be made in describing data sets and algorithms, however. Algorithms (or specifically implementations) will have to be audited by a human auditor, as automatic software verification is unlikely to be feasible in daily practice. The system should provide digital signature verification to support this however, as demonstrated before by Cushing et al. [16]. The semantic description can also be used as a machine-readable base for auditing network configuration and performance, in order to ensure accountability of the infrastructure providers. Complex constraints can be verified by querying or theorem proving. Finally, provenance recording can be done for after-the-fact auditing, as well as

for reproducibility in a scientific context.

With respect to practicality, automatic request validation as demonstrated here is a necessity for providing a timely response to user requests. The present implementation could be improved upon by support for more archetypes (as in [16]) and more complex workflows, and by leaving more of the details of where and how to execute to the system, rather than the user. This needs more advanced algorithms for matching and scheduling, however, which we plan to develop. From the system administration perspective, RDF's flexibility allows putting fewer constraints on users. Our approach allows for the translation of a simple human-understandable concept like the archetype presented in Fig. 4, into a machine-understandable concept like the model presented in Fig. 8, removing the burden of the translation from the system's administrator.

VII. Related work

One of the first work in semantic policy management is proposed by Uszok *et al.* in [17]. KAOS is composed of two core ontologies: the actor ontology, which describes people and software subjects of an action, and the action ontology, which provides support for describing actions and related context. There are four types of policies: positive or negative authorization and positive or negative obligations. This model seems to be no more in use.

The Legal Knowledge Interchange Format (LKIF) presented by Hoekstra *et al.* [18] includes a legal core ontology and a legal rule language that can be used to deploy comprehensive legal knowledge management solutions. This model lacks proper representation of the temporal aspects. Gandon *et al.* present in [19] an extension of LegalRuleML [20] for deontic reasoning on normative requirements and rules. LegalRuleML is a rule interchange language proposed by OASIS, based on RuleML (Rule Markup Language). RuleML is a unifying system of families of languages for Web rules specified through schema languages for Web documents and data. These models allow for very specific logic reasoning, which is not required by our model. We prefer to keep the modeling lighter, to allow for more flexibility.

XACML (eXtensible Access Control Markup Language) [21], is an OASIS industry standard language for access control requests and policies. It provides a common ground regarding terminology and workflow between multiple vendors building implementations of access control using XACML and interoperability between the implementations. This model is a general policy language model, while ORDL focuses on modeling digital rights over assets [22].

L4LOD is a lightweight vocabulary for expressing licensing terms in Linked Open Data [23]. Its aim is to provide the means to represent existing licensing models in RDF. However, we are using ODRL which already provides support for RDF representation of the contracts.

Palmirani *et al.* [24] introduce one of the first GDPR inspired ontologies integrated with deontic logic model, called PrOnto. PrOnto allows for privacy and data protection regulation in order to define the legal concepts in legal frameworks and the relationships among them. In the context of medical data privacy, Li and Samavi propose Data Sharing Agreement Privacy Ontology (DSAP) [25]. This ontology is specific for the medical domain, and it is not widely applicable. Our work focuses on sharing datasets, and these models do not allow for modeling the business aspects, e.g. sharing data.

VIII. Conclusion

In this paper, we present a semantic model to represent DDM policies, instrumenting an example of data sharing in the business domain. In general, data sharing is a key topic in the eScience community. Data sharing has the potential to benefit any party involved in the process and our framework can be used as an open data sharing platform in any sector.

For instance, in the hospitality domain, sharing data could be used to measure the penetration of Airbnb, and help municipalities to regulate better the phenomenon [26]. A sharing data platform which guarantees controlled access to the data as ours could be of great help to this objective.

As future work, we aim at extending the model to cover more complex workflows and policies and study the necessity of extending the ontology. Additionally, we will implement all the security steps mentioned in section VI: mainly security for users' access to the DDM, and put in place a procedure to make datasets and algorithms available in the DDM reliable. Moreover, we will complete the matching algorithm to be sure that it can deal with all of the possible policies.

References

[1] G. Stefansson, "Business-to-business data sharing: A source for integration of supply chains," *J. Production Economics*, vol. 75, no. 1-2, pp. 135–146, 2002.

[2] D. Harris, L. Khan, R. Paul, and B. Thuraisingham, "Standards for secure data sharing across organizations," *Computer Standards and Interfaces*, vol. 29, no. 1, pp. 86–96, 2007.

[3] S. van den Braak, S. Choenni, R. Meijer, and A. Zuiderwijk, "Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector," in *Proc. Conf. on Digital Government Research*. New York, NY, USA: ACM, 2012, pp. 135–144.

[4] L. Gommans, J. Vollbrecht, B. G. de Bruijn, and C. de Laat, "The service provider group framework: A framework for arranging trust and power to facilitate authorization of network services," *Future Generation Computer Systems*, vol. 45, pp. 176–192, 2015.

[5] "ODRL Information Model 2.2," https://www.w3.org/TR/odrl-model/, 2019, [Online; accessed 1-April-2019].

[6] "DL4LD," https://www.dl4ld.nl, 2019, [Online; accessed 26-April-2019].

[7] "Enabling Personal Intervention," https://delaat.net/epi/, 2019, [Online; accessed June-2019].

[8] "Green Village Sharing Platform," https://bit.ly/2HQDUNu, 2019, [Online; accessed 26-April-2019].

[9] "SURFSARA," https://bit.ly/2Z4mRNH, 2019, [Online; accessed 26-April-2019].

[10] "NLIP," https://www.nlip.org, 2019, [Online; accessed 26-April-2019].

[11] "ISHARE project," https://www.ishareworks.org/project, 2019, [Online; accessed 26-April-2019].

[12] R. Kahn and R. Wilensky, "A framework for distributed digital object services," *International Journal on Digital Libraries*, vol. 6, no. 2, pp. 115–123, Apr 2006. [Online]. Available: https://doi.org/10.1007/s00799-005-0128-x

[13] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations," *The International Journal of High Performance Computing Applications*, vol. 15, no. 3, pp. 200–222, 2001. [Online]. Available: https://doi.org/10.1177/109434200101500302

[14] M. Coppola, Y. Jgou, B. Matthews, C. Morin, L. P. Prieto, . D. Snchez, E. Y. Yang, and H. Yu, "Virtual organization support within a grid-wide operating system," *IEEE Internet Computing*, vol. 12, no. 2, pp. 20–28, March 2008.

[15] M. M. Mello, J. K. Francer, M. Wilenzick, P. Teden, B. E. Bierer, and M. Barnes, "Preparing for responsible sharing of clinical trial data," *New England Journal of Medicine*, vol. 369, no. 17, pp. 1651–1658, 2013, pMID: 24144394. [Online]. Available: https://doi.org/10.1056/NEJMhle1309073

[16] R. Cushing, L. Zhang, Y. Demchenko, C. de Laat, and P. Grosso, "Data harbours: Computing archetypes for digital marketplaces," in *Proc. Conf. International Conference on High Performance Computing and Simulation (HPCS 2019)*, 2018.

[17] A. Uszok, J. M. Bradshaw, and R. Jeffers, "Kaos: A policy and domain services framework for grid computing and semantic web services," in *Trust Management*, C. Jensen, S. Poslad, and T. Dimitrakos, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 16–26.

[18] R. Hoekstra, J. Breuker, M. D. Bello, and A. Boer, "The lkif core ontology of basic legal concepts," in *In Legal Ontologies and Artificial Intelligence Techniques*. Ceur.org, 2007.

[19] S. V. Fabien Gandon, Guido Governatori, "Normative requirements as linked data," in *JURIX 2017 - The 30th international conference on Legal Knowledge and Information Systems*, Luxembourg, Luxembourg, 2017, pp. 1–10.

[20] H.-P. Lam, M. Hashmi, and B. Scofield, "Enabling reasoning with legalruleml," in *Rule Technologies. Research, Tools, and Applications*, J. J. Alferes, L. Bertossi, G. Governatori, P. Fodor, and D. Roman, Eds. Cham: Springer International Publishing, 2016, pp. 241–257.

[21] D. Ferraiolo, R. Chandramouli, R. Kuhn, and V. Hu, "Extensible access control markup language (xacml) and next generation access control (ngac)," in *Proc. ACM Workshop on Attribute Based Access Control*, ser. ABAC '16. New York, NY, USA: ACM, 2016, pp. 13–24. [Online]. Available: http://doi.acm.org.vu-nl.idm.oclc.org/10.1145/2875491.2875496

[22] X. Maroñas, E. Rodriguez, and J. Delgado, "An architecture for the interoperability between rights expression languages based on xacml," in *Proc. Workshop for technical, economic and legal aspects of business models for virtual goods incorporation the 5th international ODRL workshop*, Sep 2009, pp. 29–47. [Online]. Available: http://www.virtualgoods.org/2009/29_VirtualGoods2009Book.pdf

[23] F. G. Serena Villata, "L4LOD Vocabulary Specification 0.2," https://ns.inria.fr/l4lod/v2/l4lod_v2.html, 2013, [Online; accessed 4-June-2019].

[24] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, and L. Robaldo, "Pronto: Privacy ontology for legal reasoning," in *Electronic Government and the Information Systems Perspective*, A. Kő and E. Francesconi, Eds. Cham: Springer International Publishing, 2018, pp. 139–152.

[25] M. Li and R. Samavi, "Dsap: Data sharing agreement privacy ontology," in *Proc. Conf. Semantic Web Applications and Tools for Life Sciences (SWAT4HCLS 2018)*. CEUR.org, 2018.

[26] G. Quattrone, A. Greatorex, D. Quercia, L. Capra, and M. Musolesi, "Analyzing and predicting the spatial penetration of airbnb in u.s. cities," *EPJ Data Science*, vol. 7, no. 1, p. 31, Sep 2018. [Online]. Available: https://doi.org/10.1140/epjds/s13688-018-0156-6

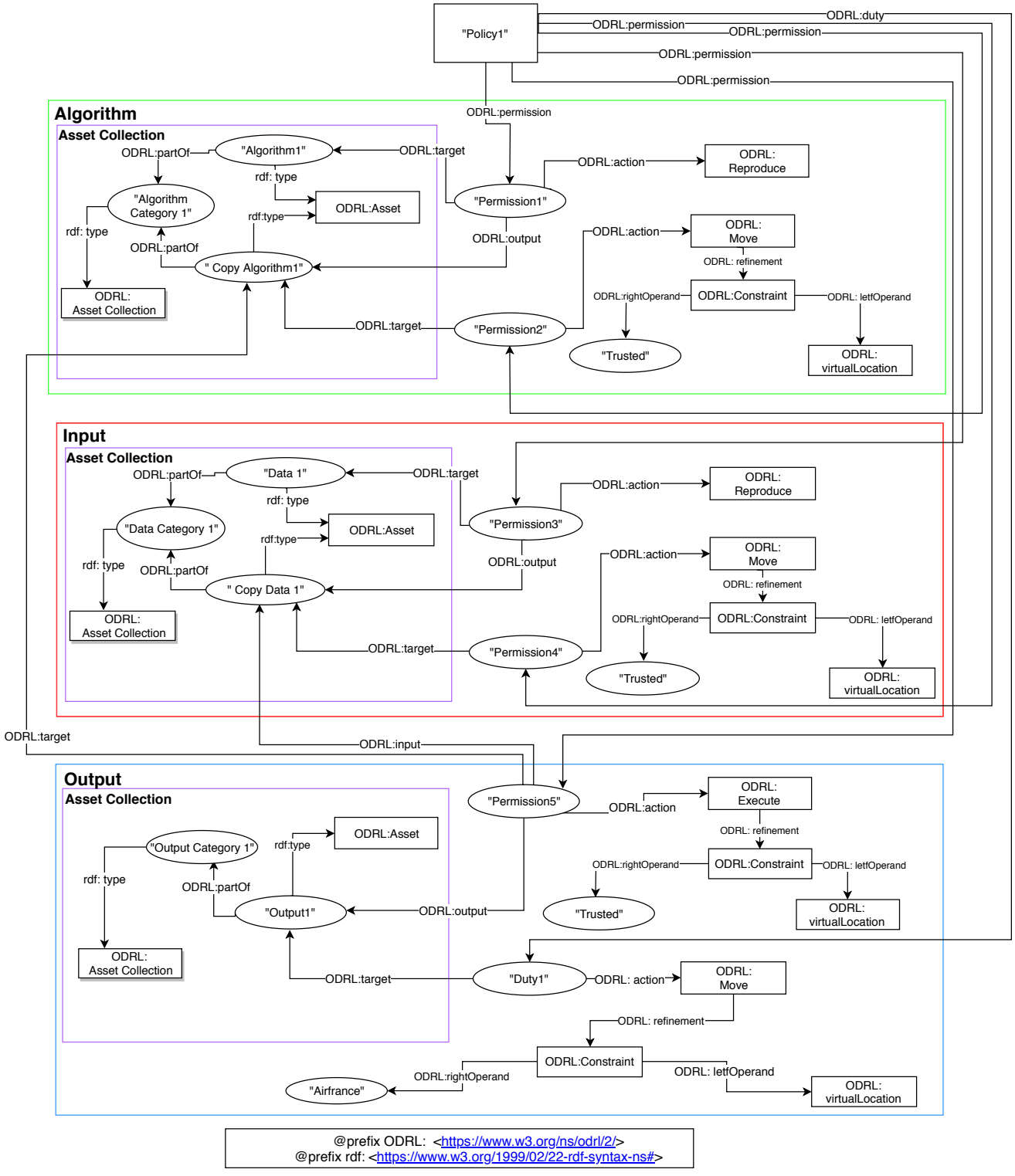

Fig. 8. An example of a policy. The figure translates the archetype shown in Fig. 4 into our semantic model.