

# Modeling and Matching Digital Data Marketplace Policies

Sara Shakeri, Valentina Maccatrozzo, Lourens Veen, Rena Bakhshi, Leon  
Gommans, Cees de Laat and Paola Grosso

# Harvard Business Review



Harvard Business Review

ECONOMY

## Managing Our Hub Economy


by Marco Iansiti and Karim R. Lakhani

FROM THE SEPTEMBER–OCTOBER 2017 ISSUE

WHAT TO READ NEXT

The IT Transformation Health Care Needs

SUMMARY SAVE SHARE COMMENT 3 TEXT SIZE PRINT \$8.95 BUY COPIES



THOMAS M. SCHEER/EYEEM/GETTY IMAGES

### I. The Problem

The global economy is coalescing around a few digital superpowers. We see unmistakable evidence that a winner-take-all world is emerging in which a small number of “hub firms”—including Alibaba, Alphabet/Google, Amazon, Apple, Baidu, Facebook, Microsoft, and Tencent—occupy central positions. While creating real value for users, these companies are also capturing a disproportionate and expanding share of the value, and that’s shaping our collective economic future. The very same technologies that promised to democratize business are now threatening to make it more monopolistic.

Data value creation  
monopolies



Create an equal  
playing field



Sound Market  
principles

<https://hbr.org/2017/09/managing-our-hub-economy>

# Main problem statement

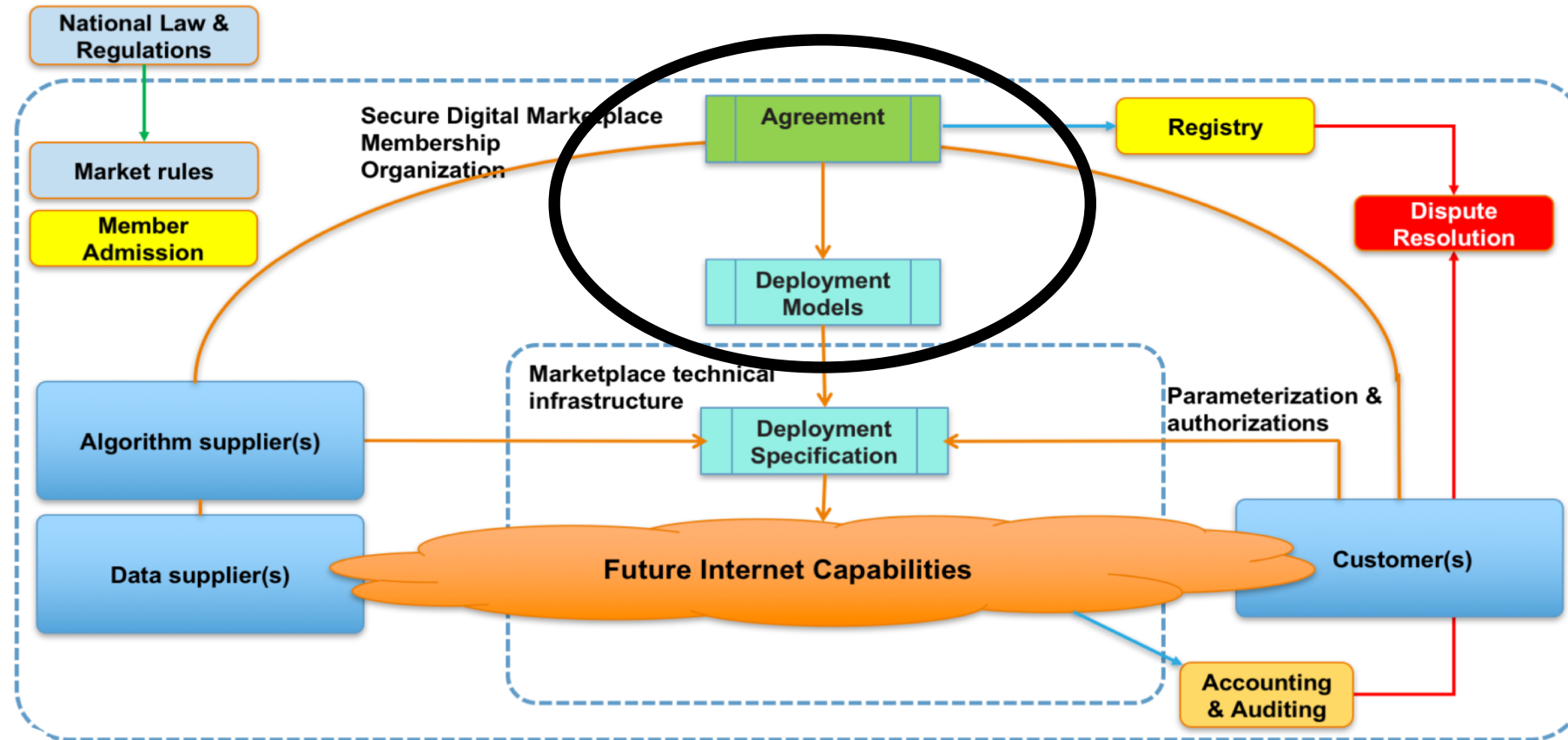
- There is lots of data out there that is not shared (99%)
- FAIR is typically not fair ;-), but limited by policy and/or law
  - the A in FAIR is about access, trust is hard to implement across domains
- Organizations that normally compete have to bring data together to achieve a common goal/benefit!
- The shared data may be used for that goal but not for any other!
- Expected use is fine but unexpected use/mission creep...
- Data processed by alien algorithms in foreign data centers... Hmmm...
  - How to organize data processing alliances?
  - How to enforce policy using modern Cyber Infrastructure?
  - How to translate law policy from strategic via tactical to operational level?
  - What are the different fundamental data infrastructure models to consider?

# Approach

- Strategic:
  - Translate legislation into machine readable policy
  - Define data use policy
  - Trust evaluation models & metrics
- Tactical:
  - Map app given rules & policy & data and resources
  - Bring computing and data to (un)trusted third party
  - Resilience
- Operational:
  - TPM & Encryption schemes to protect & sign
  - Policy evaluation & docker implementations
  - Use VM and SDI/SDN technology to enforce
  - Block chain to record what happened (after the fact!)



# Digital data marketplace framework



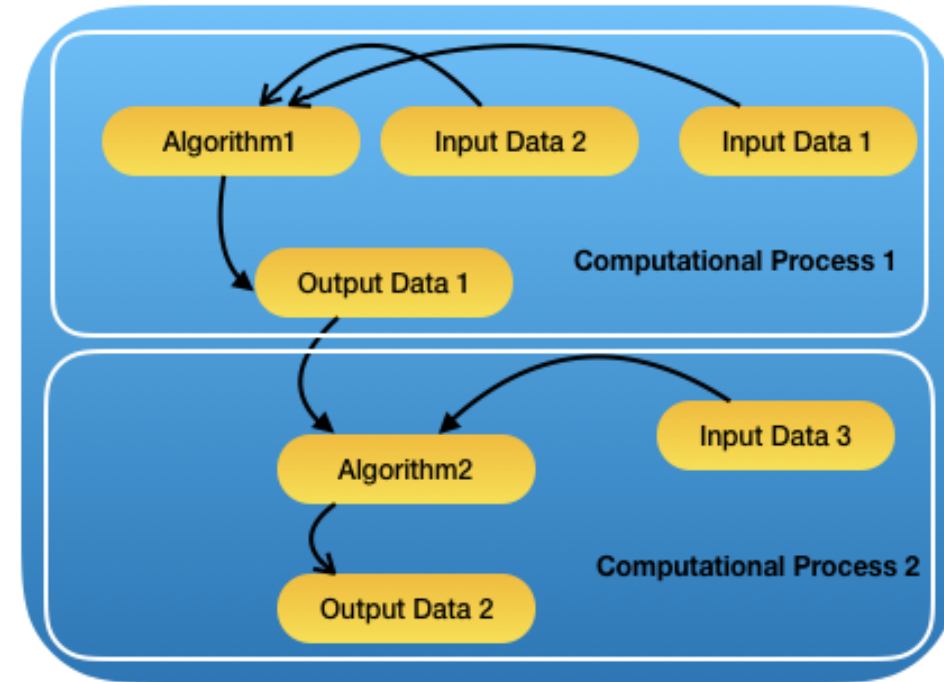
<https://dl4ld.nl/>

# Goals

- Use semantic modelling to represent data sharing policies agreed between partners in a DDM
- Demonstrate checking a user request against the usage policies

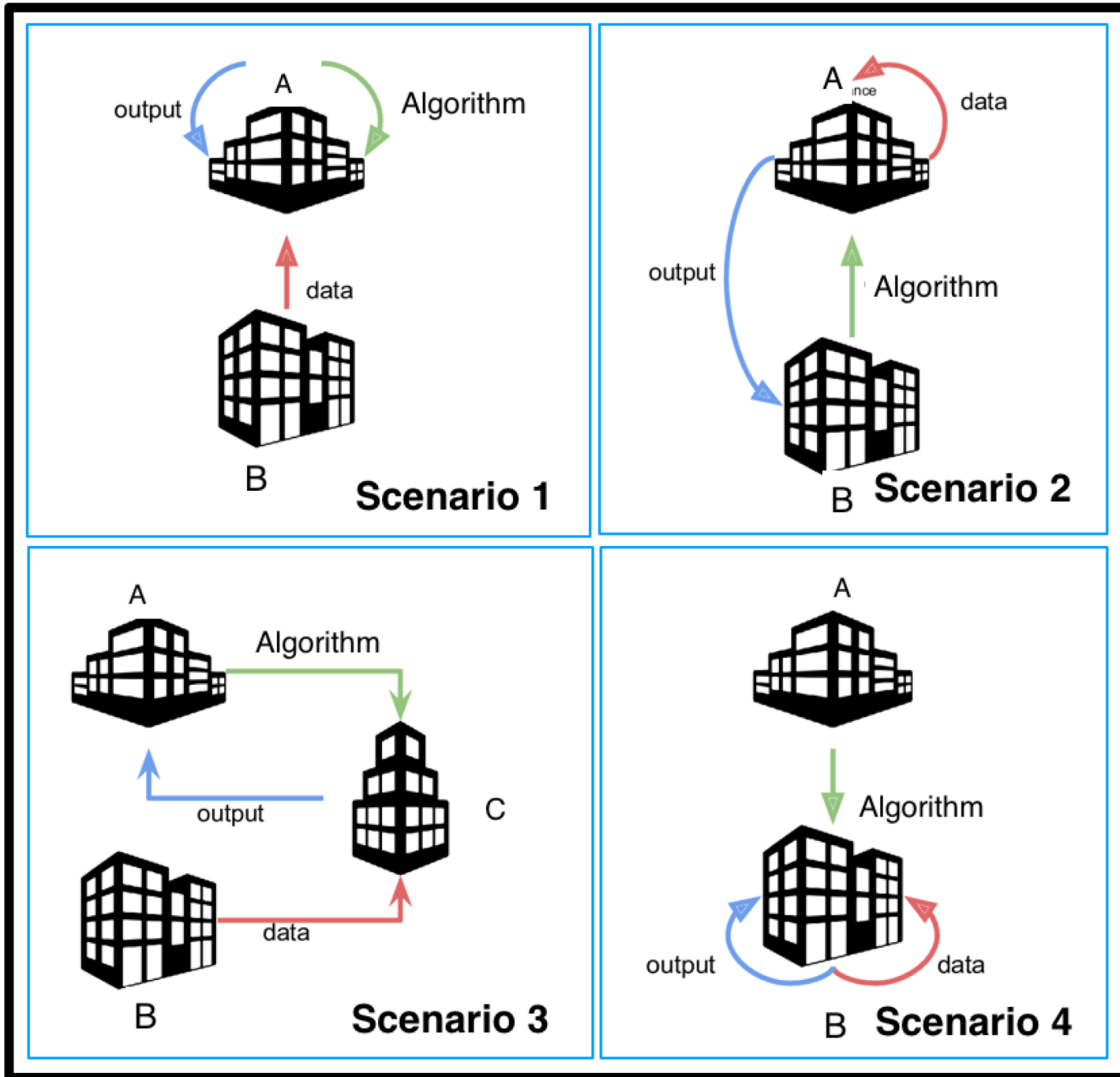
# DDM Application

- Two kinds of resources can be shared in the proposed DDM system
  - Algorithm
  - Data
    - Input Data
    - Output Data



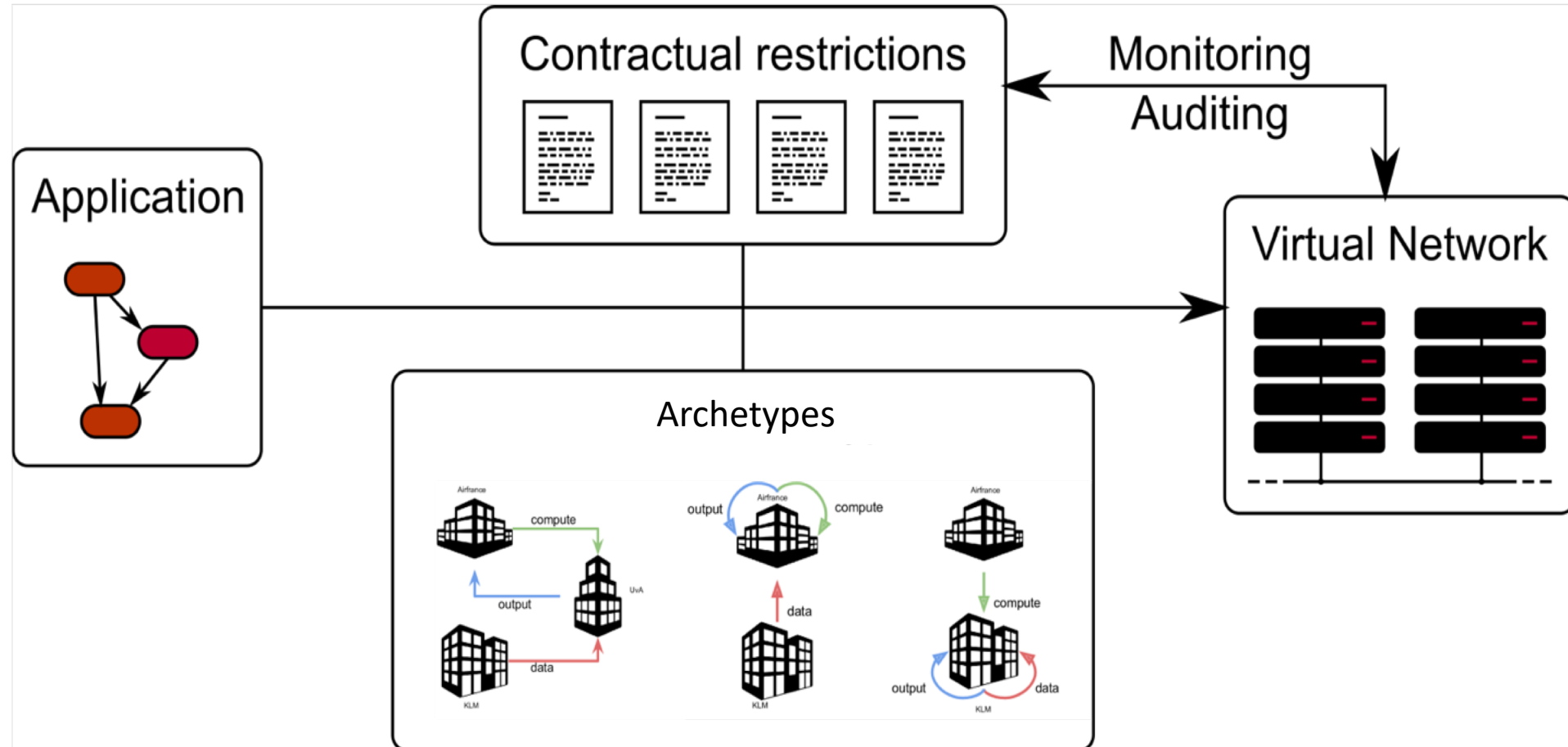
# DDM Archetype

- A Scenario that determines the permitted transmissions of the shared digital resources.





# Request Handling in a DDM

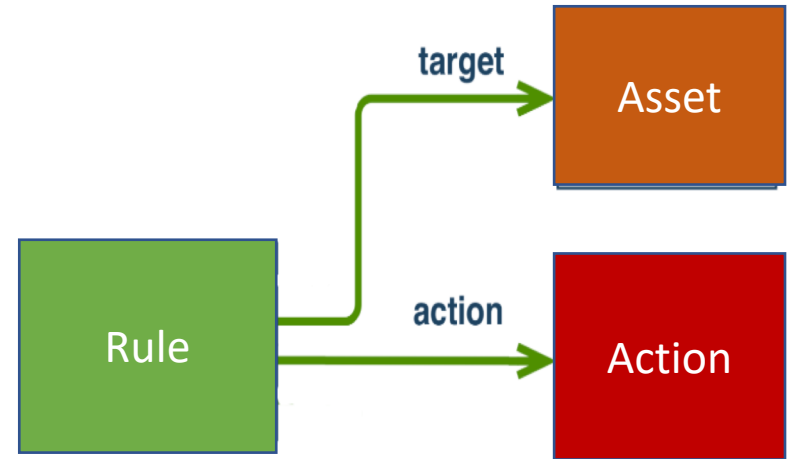


# Semantic Model Requirements

- Describe how resources can be shared and used by different parties
- Required permissions to support archetypes
  - Copying the asset to a particular location
  - Moving the asset to a particular location
  - Execution on a particular location
  - Moving the results of the whole operation (output) to a particular location

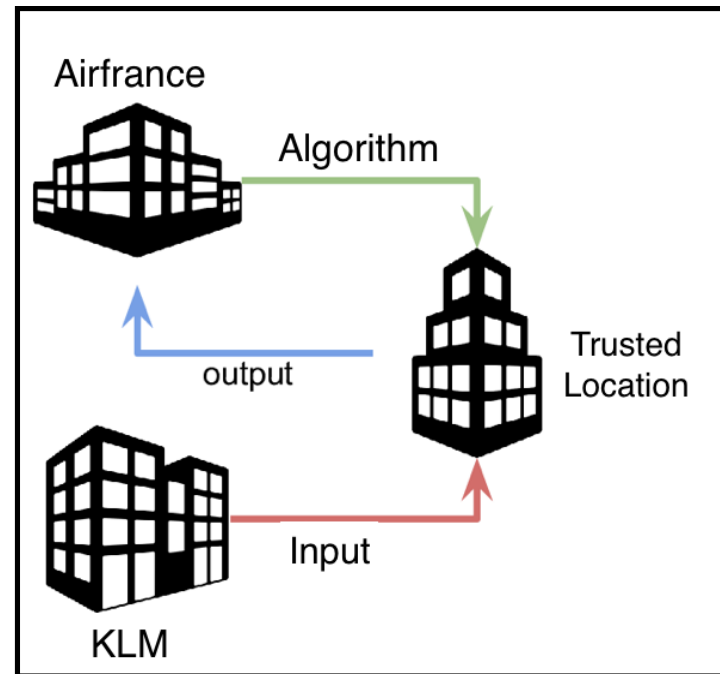
# ODRL : Open Digital Rights Language

- An ontology designed to model permissions, obligations, and prohibitions concerning digital resources.
- The main classes are:
  - Asset: a digital resource, e.g., data or algorithms
  - Action: an activity performed on an Asset
  - Rule: constrains an Action performed on an Asset.

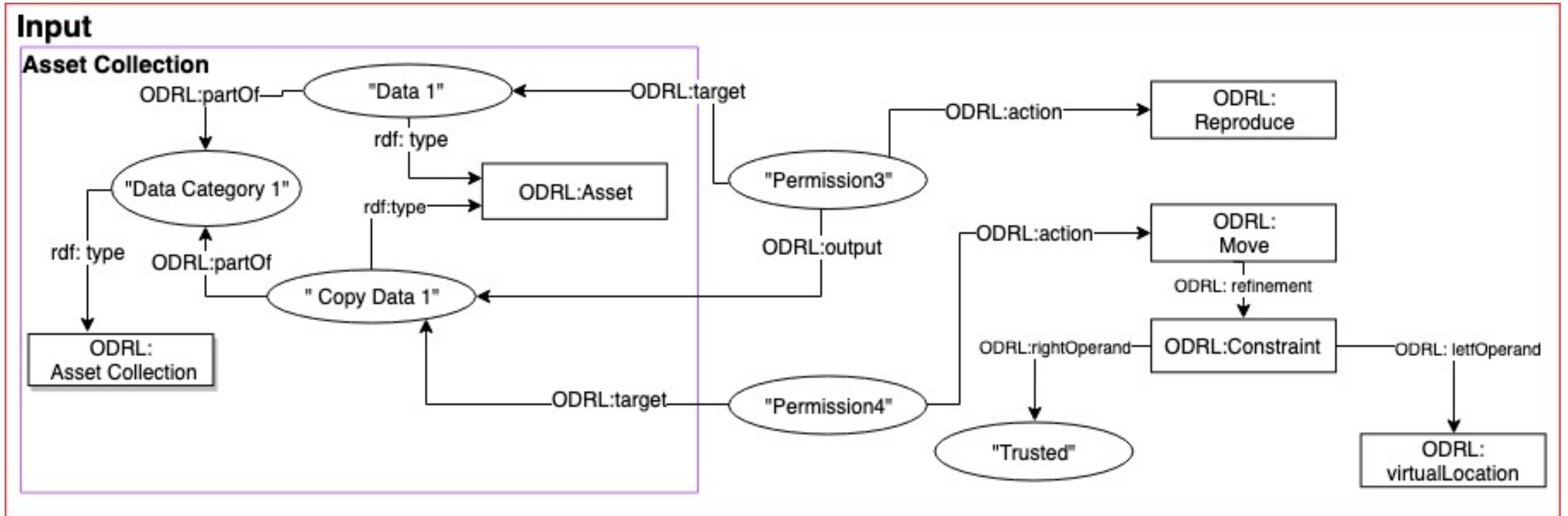


[<https://www.w3.org/TR/odrl-model/>]

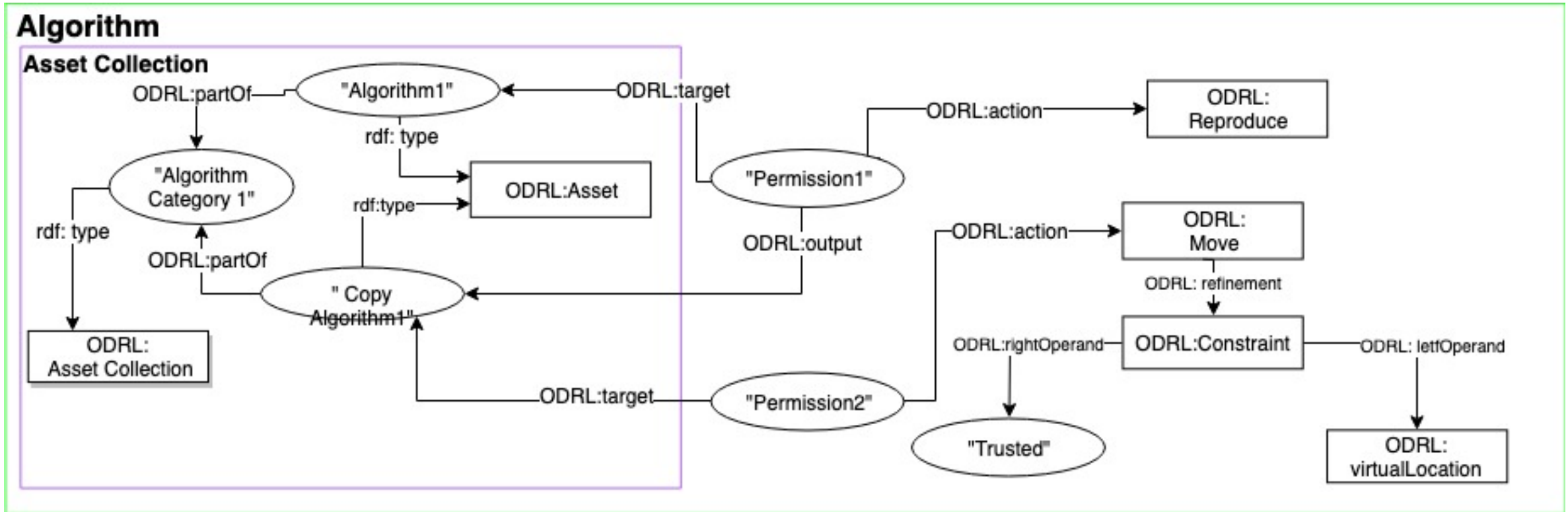
# Example archetype



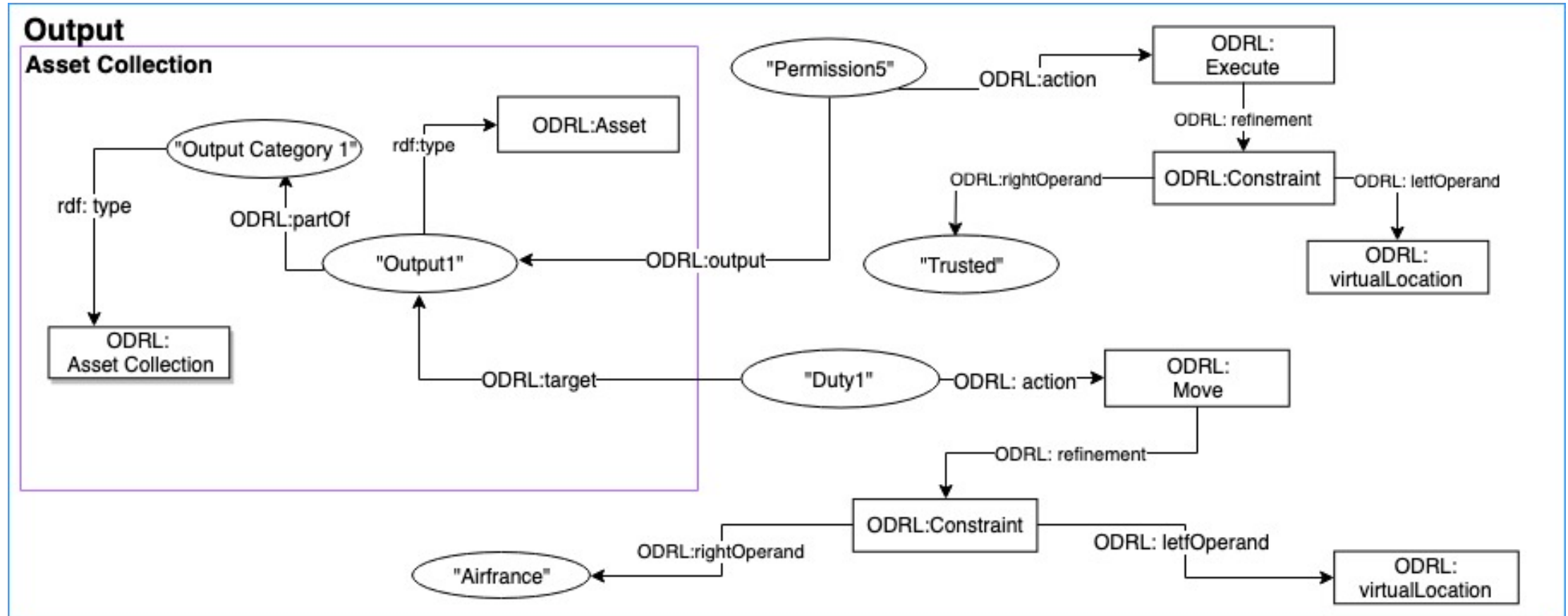
# Permissions for input



# Permissions for algorithm



# Permissions for output



# Matching Module

- Automatic management of user request:
  - Users can submit a request to use specific datasets or algorithms, specifying the location of execution.
- The request must be matched with the available archetypes in DDM.
- Matching module verifies whether the request is permitted and approve or reject it.



# Matching algorithm

---

**Alg 1: Matching Algorithm**

---

**Input:** Algorithm1, Data1, Execution Location, Output Location

```
1 Function moveAllowed(Asset, Location)
2   | if location(Asset) ≠ 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
```

---

# Discussion

- The system must provide sufficiently broad access
  - ODRL is a powerful rights description language, and the use of semantic technology makes it easy to extend the ontology if needed.
- It must ensure accountability of all parties involved
  - To ensure accountability of users, requests need to be matched against the archetypes specified in the contracts.
- It must be practicable
  - The present implementation could be improved upon by support for more archetypes and more complex workflows and more flexible matching.

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

# Summary and future work

- Enabling algorithm and data sharing in the eScience community
- Proposing a semantic model to represent DDM policies
- Our framework is an essential component in DDMs
- Future work
  - Extending the model to cover more complex workflows and policies
  - Extending the matching algorithm to be sure that it can deal with all of the possible policies and select the best
  - User interface to guide user towards a permitted request

<https://www.esciencecenter.nl/>