

Digital Data Markets: real time ICT for logistics

Data Logistics 4 Logistics Data (dl4ld)

Research

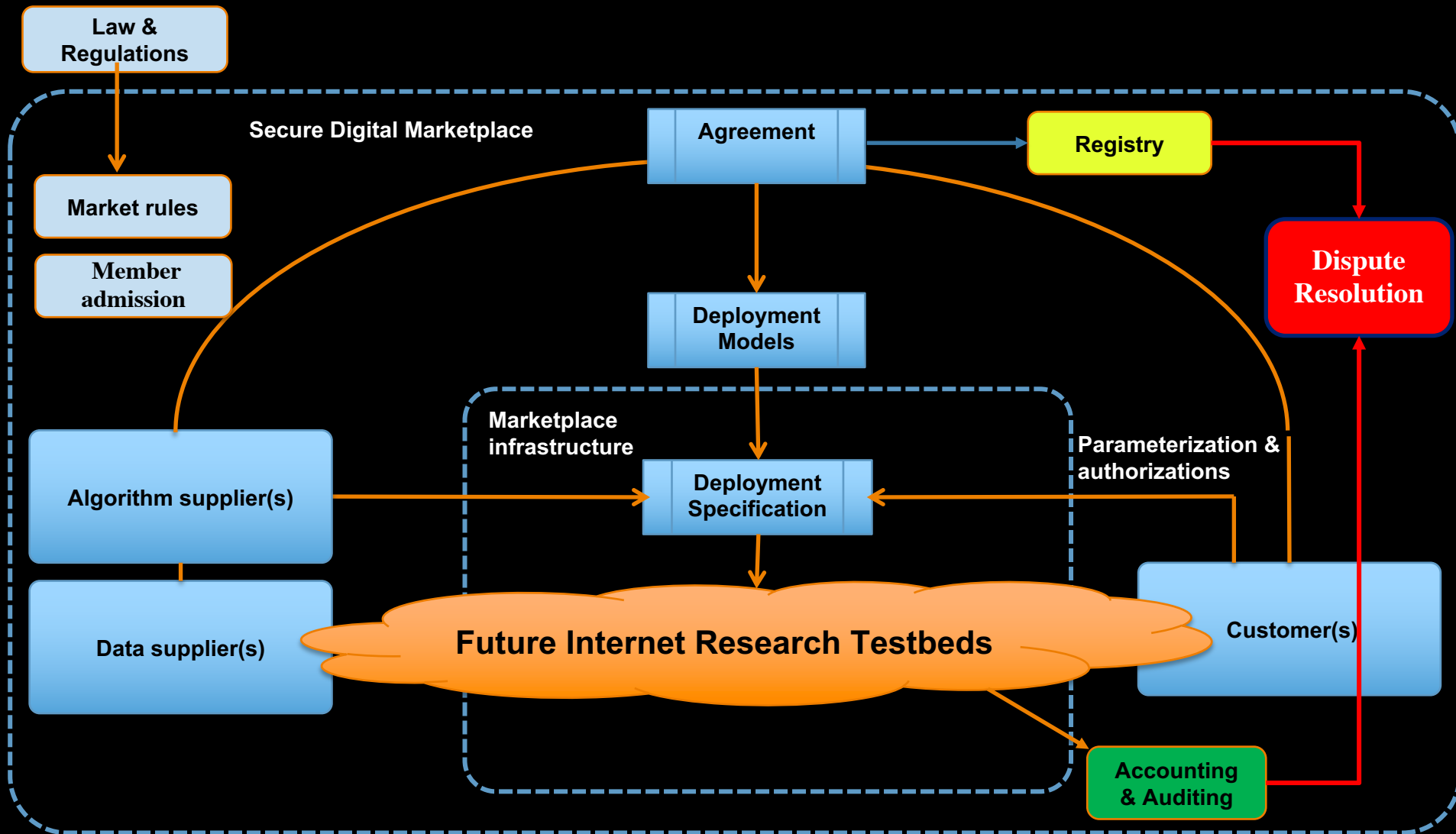
Cees de Laat



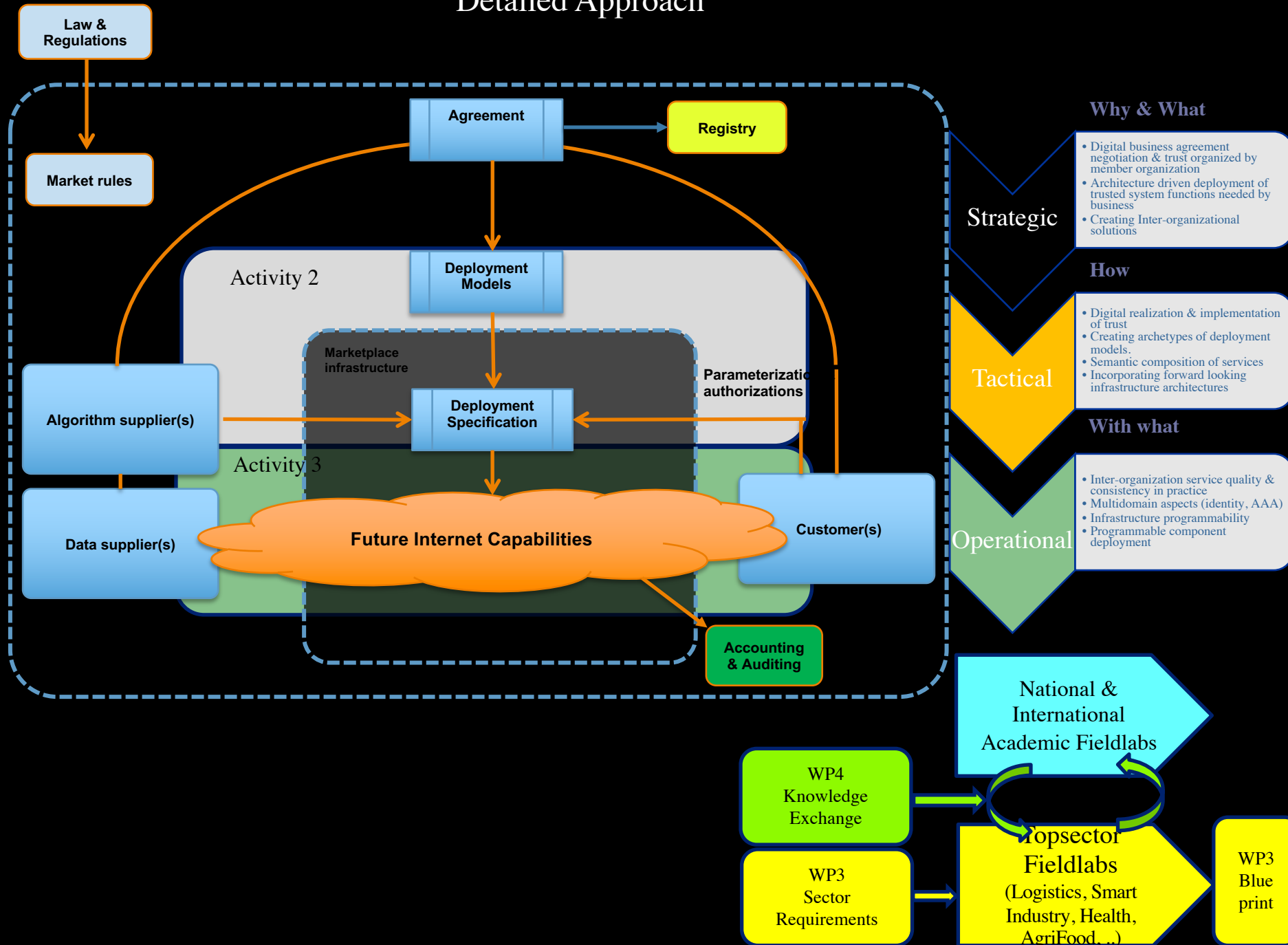
Main problem statement

- Organizations that normally compete have to bring data together to achieve a common goal!
- The shared data may be used for that goal but not for any other!
- Data may have to be processed in untrusted data centers.
 - How to enforce that using modern Cyber Infrastructure?
 - How to organize such alliances?
 - How to translate from strategic via tactical to operational level?
 - What are the different fundamental data infrastructure models to consider?

Secure Digital Market Place Research



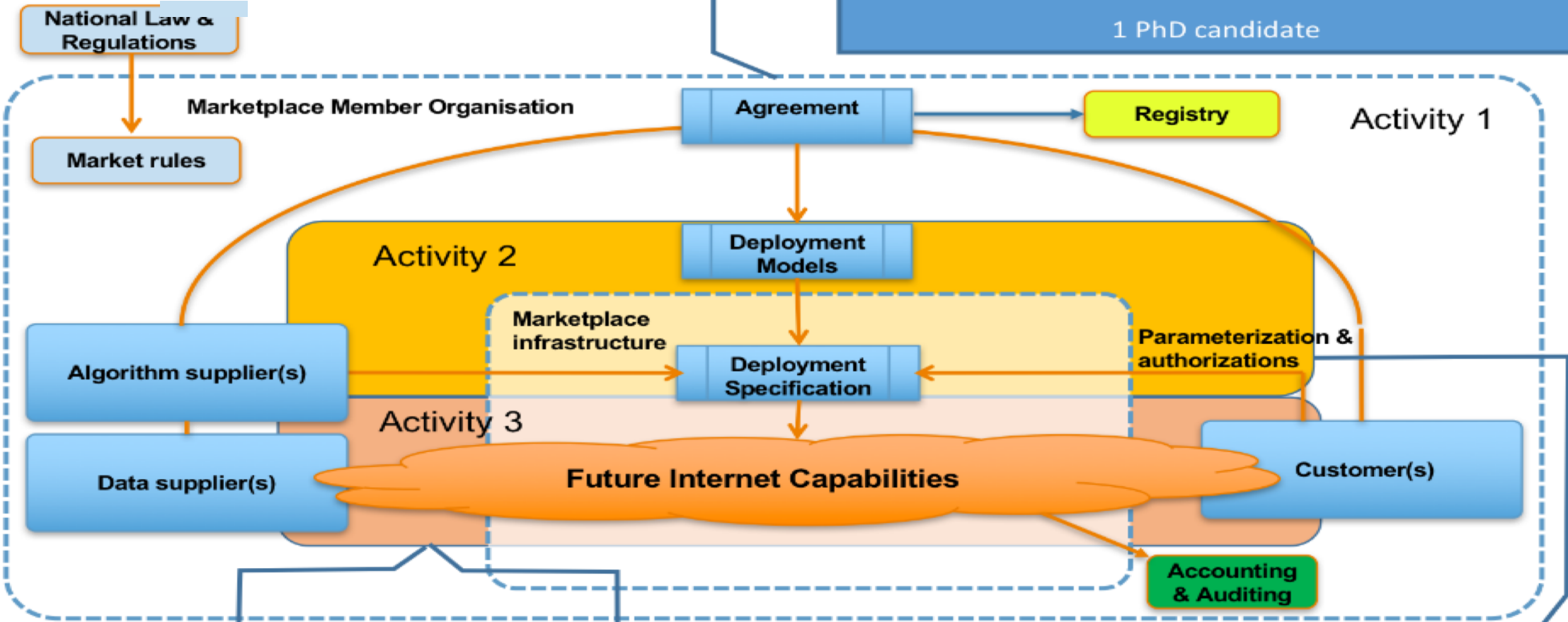
Detailed Approach



WP2 Research activity layout and staff involvement

CDL: Cees de Laat
 TVE: Tom van Engers
 SK: Sander Klous
 PG: Paola Grosso
 LG: Leon Gommans

TVE: Digital business agreement negotiation & trust.
 LG: Architecture driven deployment of trusted systems
 SK: Inter-organizational solution development
 1 PhD candidate



Amsterdam Datahub
 SK: Inter-organization service quality & consistency in practice
 1 PhD candidate

Openlab, KLM, Ciena, GLIF
 CDL: Multidomain aspects
 PG: Infra programmability
 LG: Programmable component deployment
 1 Prgmr
 1 PhD candidate

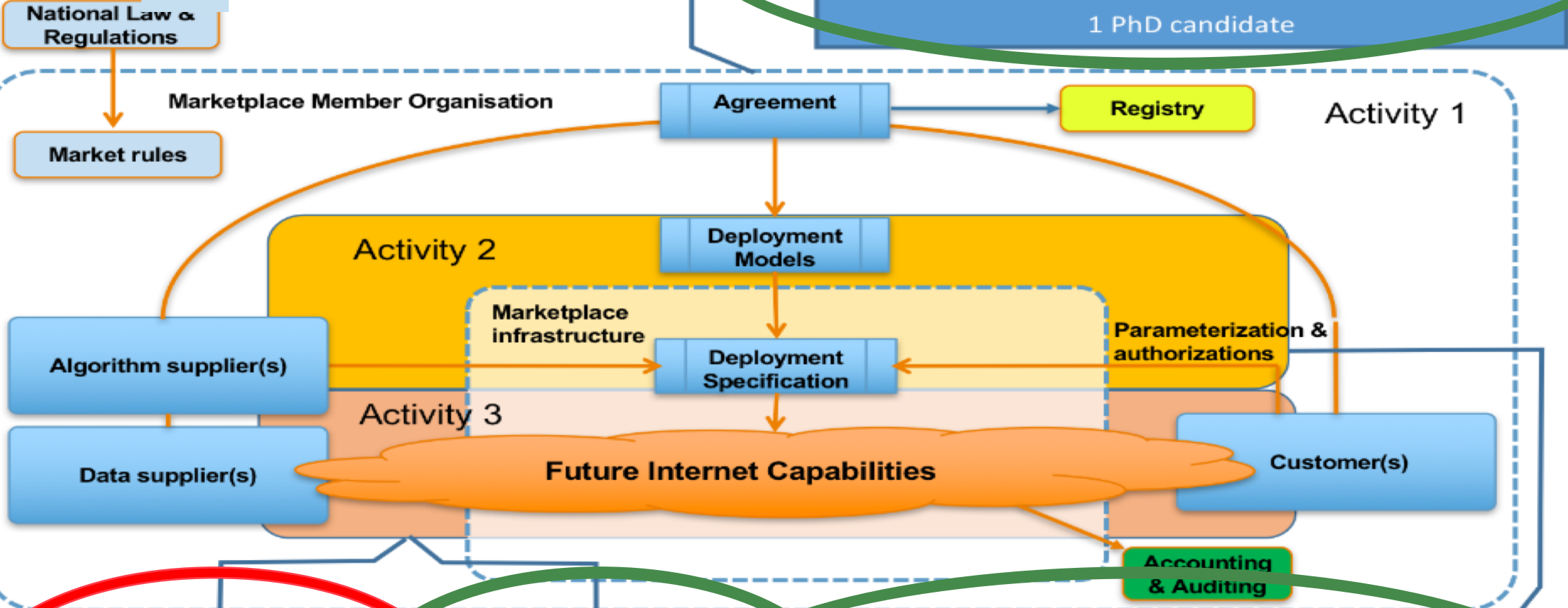
TVE: Digital realization of trust
 LG: Creating archetypes of deployment models.
 PG: Semantic composition of services
 CDL: Forward looking architectures
 1 Postdoc researcher



WP2 Research activity layout and staff involvement

CDL: Cees de Laat
 TVE: Tom van Engers
 SK: Sander Klous
 PG: Paola Grosso
 LG: Leon Gommans

TVE: Digital business agreement negotiation & trust.
 LG: Architecture driven deployment of trusted systems
 SK: Inter-organizational solution development
 1 PhD candidate



Amsterdam Datahub
 SK: Inter-organization service quality & consistency in practice
 1 PhD candidate

Openlab, KLM, Ciena, GLUE
 CDL: Multidomain aspects
 PG: Infra programmability
 LG: Programmable component deployment
 1 PhD candidate

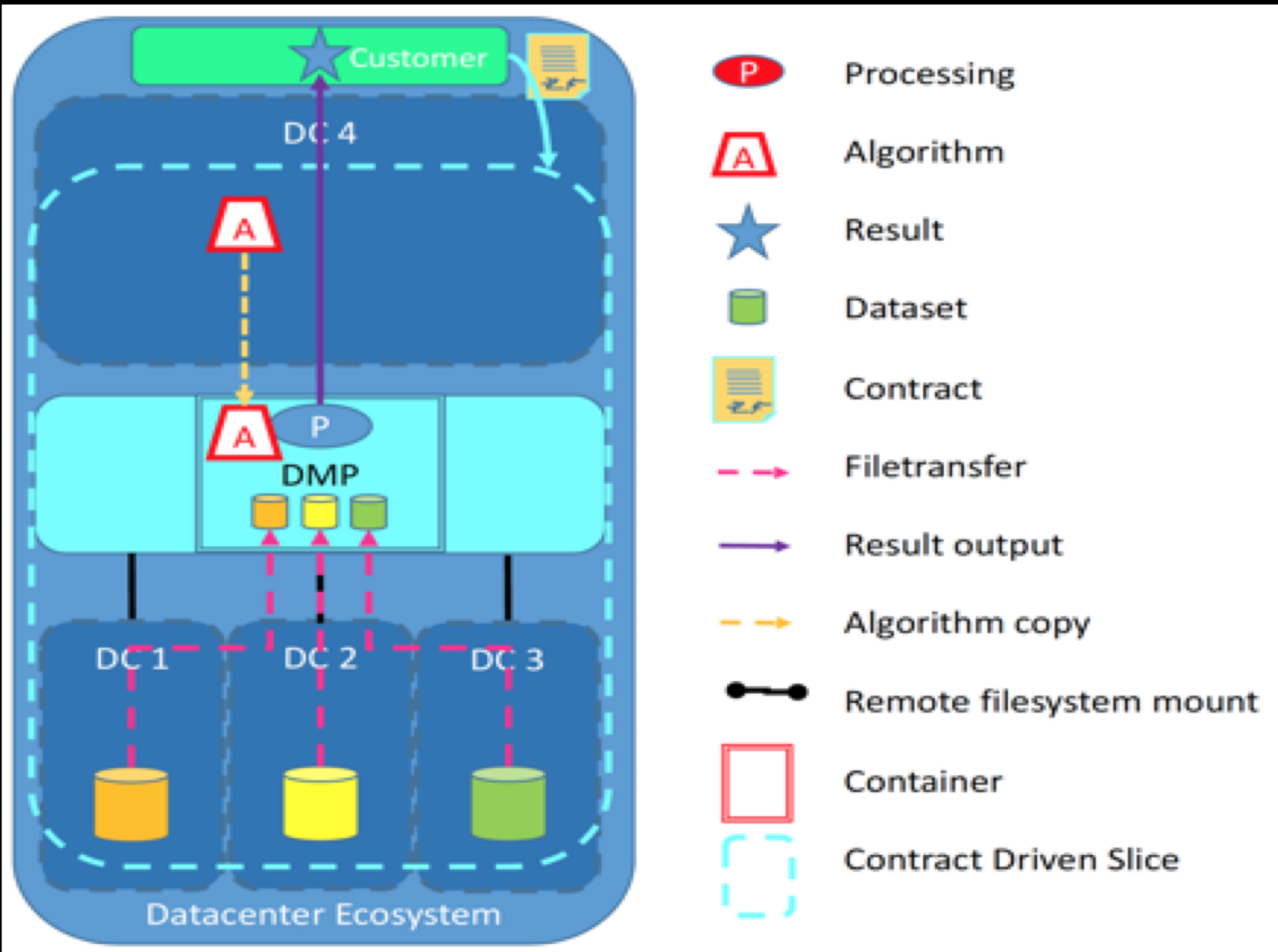
TVE: Digital realization of trust
 LG: Creating archetypes of deployment models.
 PG: Semantic composition of services
 CDL: Forward looking architectures
 1 Postdoc researcher



What have we been doing?

- Studying and defining draft Policy
- Working out some Archetypes
- Implementing a proof of concept using several distributed DTN's and dockers on kubernetes.
- Working on a demo for SC18 in Dallas TX, 11-16 Nov.
- Generic model for Archetypes
- Tactical operation of Digital Data Markets
- Optimization of degrees of freedom == value

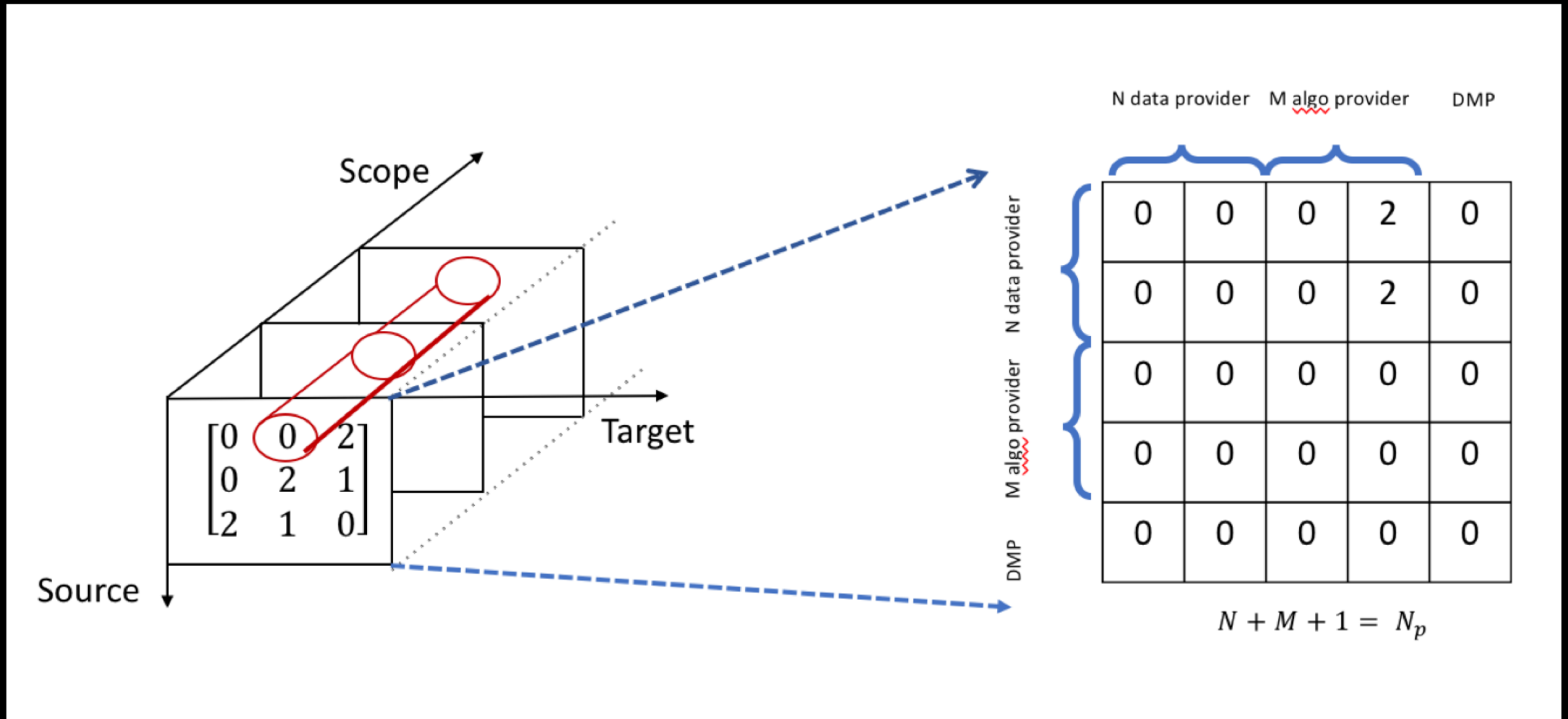
DMP archetypes and their representation



On the left one of the many collaboration models within a DMP. We call this archetype. One DMP can support multiple archetypes depending on the contracts between partners.

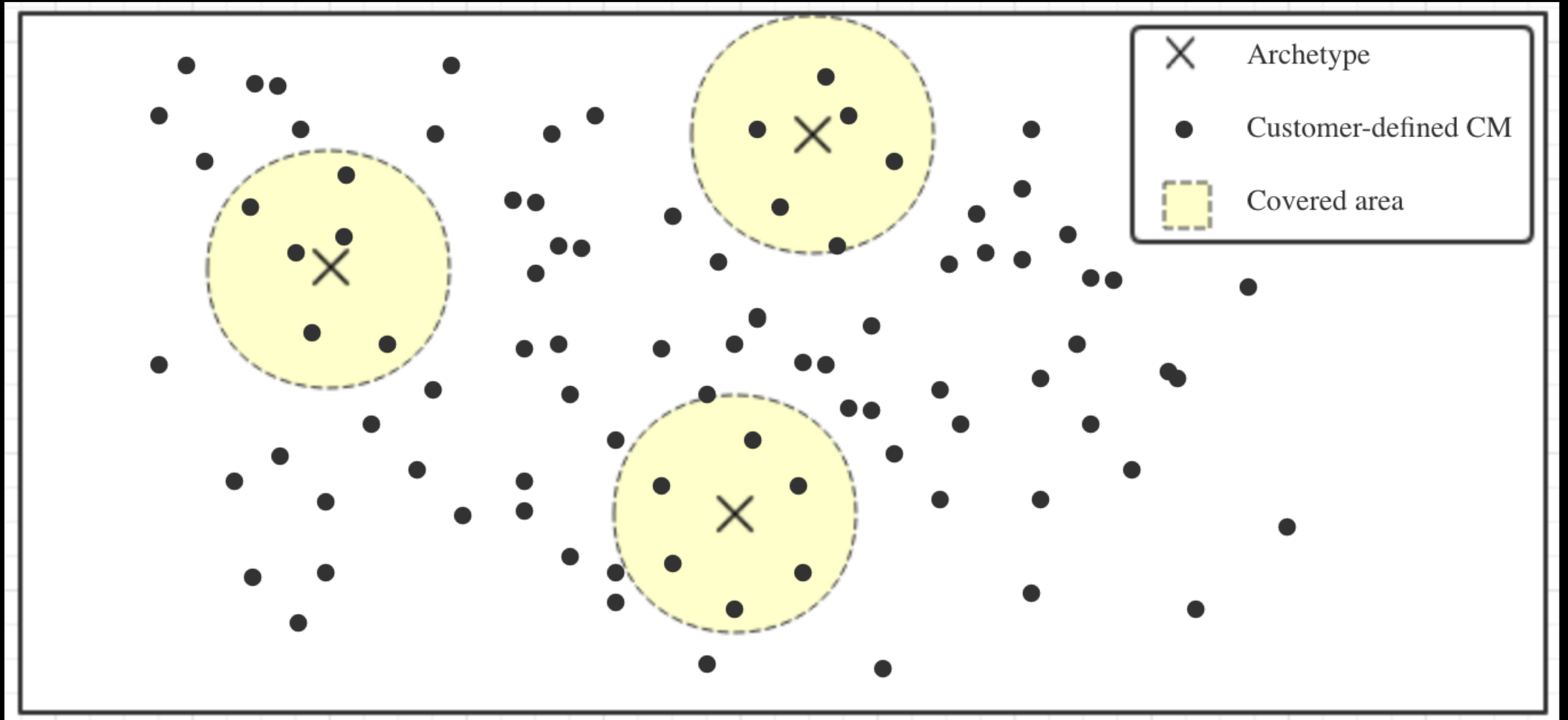
To match application/user requests to the archetype we need to model the archetype on the left in generic ways.

Models for archetypes



Parties in the DMP collaborate across a number of scopes: data, computing and output. They share data, they share algorithms and they can share results. The matrix on the left represent the level of collaboration between two parties in each of the scope. In the previous slide we had four parties plus the DMP exchange, so we have a 5x5 matrix.

Matching requests



Our work is to match a customer application to the ‘closest archetype’.

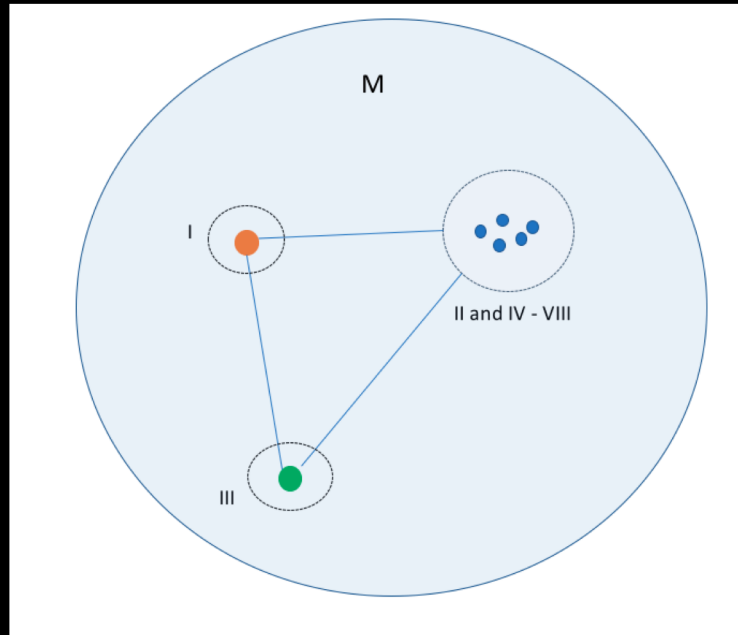
Dimensions of DMP

Coverage

How well can we satisfy users request with the available archetypes?

Extensibility

Can a DMP provide more archetypes to user?



Precision

How well the archetype database of a given DMP fits a request by customer?

Flexibility

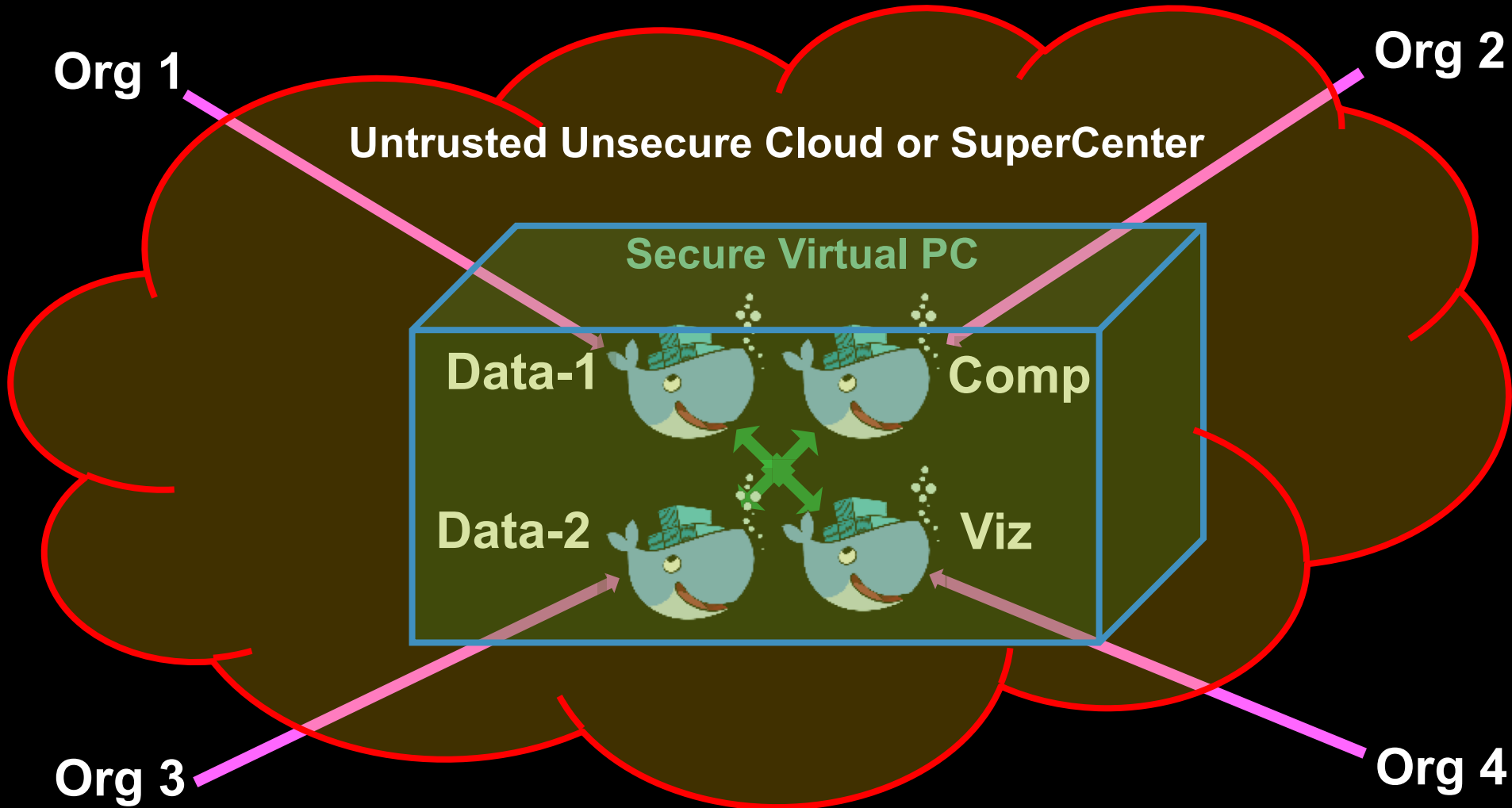
How easily the requests from potential customers could be satisfied?

We defined four metrics to determine the ‘richness’ of a DMP.

Secure Policy Enforced Data Processing



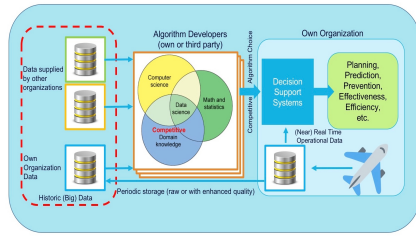
- Bringing data and processing software from competing organisations together for common goal
- Docker with encryption, policy engine, certs/keys, blockchain and secure networking
- Data Docker (virtual encrypted hard drive)
- Compute Docker (protected application, signed algorithms)
- Visualization Docker (to visualize output)



Training AI/ML models using Digital Data Marketplaces

Creating value and competition by enabling access to additional big data owned by multiple organizations in a trusted, fair and economic way

The more data - the better: an aircraft maintenance use-case



- AI/ML algorithm based Decision Support Systems create business value by supporting real-time complex decision taking such as **predicting the need for aircraft maintenance.**

- Algorithm quality increases with the availability of aircraft data.

- Multiple airlines operate the same type of aircraft.

- **Research Question:** "How can AI/ML algorithm developers be enabled to access additional data from multiple airlines?"

- **Approach:** Applying Digital Data Marketplace concepts to facilitate trusted big data sharing for a particular purpose.

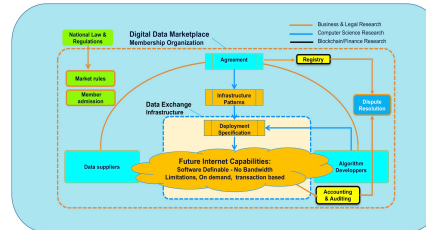
Digital Data Marketplace enabling data sharing and competition

A **Digital Data Marketplace** is a membership organization supporting a common goal: e.g. *enable data sharing to increase value and competitiveness of AI/ML algorithms.*

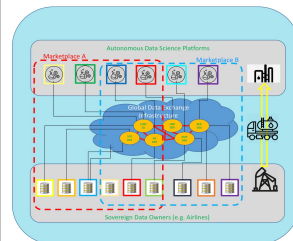
Membership organization is institutionalized to create, implement and enforce membership rules organizing **trust**.

Market members arrange **digital agreements** to exchange data for a **particular purpose** under specific conditions.

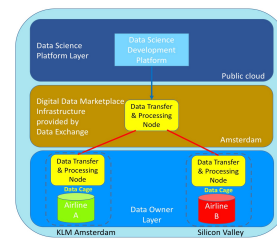
Agreements subsequently drive data science transactions creating processing infrastructures using infrastructure patterns offered by a Data Exchange as **Exchange Patterns**.



Researching Exchange Patterns to support Digital Data Marketplaces



Data Exchange Model



Research Infrastructure

Trust Modelling:
What is the optimal infrastructure archetype, describing storage and processing locations and their relationships, which best suit member requirements when considering risk?

Processing Models:
What are the implications of distributing data processing across membership organization owned infrastructures in terms of achievable model accuracy and processing performance using federated/distributed models vs centralized models?

Marketplace Reference Architecture:
What constitutes a marketplace? Researching needed functions, personas, flows, credentials, contracts & rules, conflict resolution, and much more ...

Research Elements



Leon Gommans, Anne Savelkoul, Wouter Kalfbeek, Dirk van den Herik, David Langereveld, Erik Uzermans, Floris Freeman, Brend Dikkers, Cees de Laat, Tom van Engers, Wouter Los, Paola Grosso, Joseph Hill, Reggie Cushing, Giovanni Sileno, Lu Zhang, Amesh Deltjo, Thomas Baeck, Willem Koeman, Laurie Strom, Axel Berg, Gerben van Malenstein, Kalidhar Voruganti, Rodney Wilson, Patricia Florisi

<http://sc.delaat.net>

Nov 11-16, Dallas (TX), CIENA booth 2847 SURF booth 2041

2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

Phoenix Pittsburgh Seattle Tampa Reno Austin Portland New York Salt Lake City Denver New York Austin Salt Lake City Denver Dallas Denver

No more welcome due to the EU GDPR regulation :-(

AI FRANCE KLM TRANSFIDES evofenedex ciena BIZZdesign

DATA DIGEST ORACLE THALES SCinet

SC2018 Workshop and Technical program presentations & involvement by USA group members.

Donor's:

- INDISI workshop organized by Paola Grosso, Mary Heater, Michelle Zhu and Eya Baldu.
- "Towards the Network for Data Intensive Science (INDISI)"
- Date: Sunday, November 11, 2018, Time: 09:00 - 17:30 at the Kay Bailey Hutchison Convention Center, Dallas, TX.
- Conference papers in IEEE repository.
- INDISI presentation/paper by Joseph Hill (speaker), Michael Alshorji, Paola Grosso: "Tracking network flows with PM".
- presentation
- INDISI presentation/paper by Amesh Deltjo, Tom van Engers, Leon Gommans, Cees de Laat (speaker): "Social Computational Trust Model (SC2TM): A Framework to Facilitate the Selection of Partners".
- presentation
- IACS 2018 "Big Workshop on Irregular Applications: Architectures and Algorithms" presentation by Merijn Elwies Venetrasen (speaker), Axel Luca Vandenbosch and Cees de Laat. "Mix and Match: A Model Driven Runtime Optimization Strategy for BFS on GPUs".
- presentation

Impressions:

- VLOG by Henrik Thomsen (SURF)

Digital Market Places

Data Harbours

Data Harbours: A compute infrastructure for data marketplaces

Building User-friendly Data Transfer Nodes

Building User-friendly Data Transfer Nodes

SCinet contributions from Holland

ENVI: Plus

Dynamic infrastructure planning and provisioning for time critical applications in clouds

SCinet

Team members:

- UvA:
 - J.P. Velders, teamlead Rosterheads
 - dr. Paola Grosso, teamlead
 - Sjoerd INDISI
 - Erik Konstant, Student Volunteer
 - gerd@cs.uva.nl, Ludo, 2018
- SURF:
 - Pieter de Boer BSc, Rosterheads
 - Peter Boon MSc, Rosterheads
 - Mary Heater BA, NovINDISI
- Datalys:
 - Cedric, Bash, Edge and WiFi

Contributions: INDISI Workshop

SCinet SC18 network (pdf)

SC18

SC18 booth crew

Donor abstracts

1 Data Harbours: A compute infrastructure for data marketplaces.

Reginald Cushing, Tim van Zalingen, Joseph Hill, Paola Grosso, Cees de Laat, Lu Zhang, Leon Gommans, Rodney Wilson, Marc Lyonnais, Vijayaraj Doraswamy, Parthiv Purush, Kalidhar Voruganti, Ching Wang

Sharing and computing on sensitive data is a challenge. Competing parties have many reasons to be cautious about their data yet there are situations where sharing data between competitors results in a win-win scenario whereby both parties gain an advantage. Under real world conditions, a solution that leads to be able to share more sensitive data whilst still be in control? This prototype demonstrates such a scenario where competing parties own assets (data and compute) and allow processing under strict policies which define which compute can access which data and where can processing take place. See:

- slides
- DLT-D website
- CIENA media release

2 Building User Friendly Data Transfer Nodes.

Joseph Hill

Absent: This demonstration will show how Data Transfer Nodes can be combined with existing applications to provide researchers a familiar interface with the enhanced performance of a DTN. Elicited and modified have been modified to utilize E2DN to perform event domain transfer. This allows researchers to transfer files to peers across organizations and long distances at high speed without having to interact with a DTN directly.

- slides

3 Training AI/ML models using Digital Data Marketplaces.

Leon Gommans, Anne Savelkoul, Wouter Kalfbeek, Dirk van den Herik, David Langereveld, Erik Uzermans, Floris Freeman, Brend Dikkers, Cees de Laat, Tom van Engers, Wouter Los, Paola Grosso, Joseph Hill, Reggie Cushing, Giovanni Sileno, Lu Zhang, Amesh Deltjo, Thomas Baeck, Willem Koeman, Laurie Strom, Axel Berg, Gerben van Malenstein, Kalidhar Voruganti, Rodney Wilson, Patricia Florisi

When using your own data, AI can solve problems. When sharing your data, AI can transform industries. This presentation discusses an approach called the Digital Data Marketplace, allowing companies members to share their data with other members to achieve benefits and AI to help organizations build solutions on its own. To show Digital Data Marketplaces to be used for many applications including Science, Smart Industry, Logistics, Health, Agriculture and more, we foresee the need to create Data Exchanges that offer common infrastructure capabilities. We will use an actual positive maintenance use case as an example, where an industry technology consortium could play a key role.

- presentation

4 Dynamic infrastructure planning and provisioning for time critical applications in clouds.

Sjoerd Konstantin, Paul Martin, Wang Junchao, Heun Zhou, Yang Hu, Zhiming Zhao

The Dynamic Real-time Infrastructure Planner (DRIP) is a service suite for planning and provisioning networks of virtual machines and then deploying distributed applications across those networks, respecting the critical infrastructure domain data based on specific time constraints, defined with the machine models. The DRIP contains multiple services that

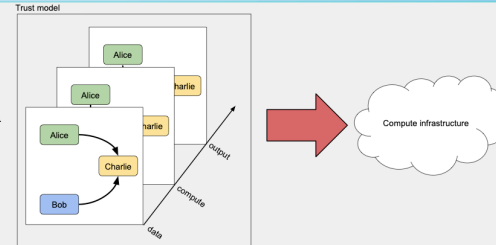


Dataharbours: computing archetypes for digital marketplaces

Reginald Cushing, Lu Zhang, Paola Grosso, Tim van Zalingen, Joseph Hill, Leon Gommans, Cees de Laat, Vijaay Doraiswamy, Purvish Purohit, Kaladhar Voruganti, Craig Waldrop, Rodney Wilson, Marc Lyounnais

The problem

How can competing parties share compute and data? The architecture of a digital marketplace is an active research field and has many components to it. Here we investigate a federated computing platform which is molded into different **archetypes** based on **trust** relationships between organizations.



The components

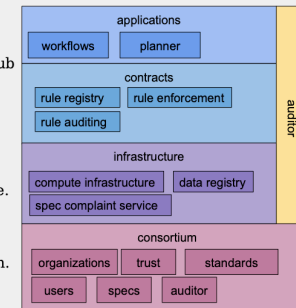
Consortium: is an initial document which brings together organizations that wish to collaborate. It defines static information such as keys to identify parties.

Infrastructure: A single domain organization infrastructure that securely hosts data, compute containers and, optionally, compute infrastructure. We dub this infrastructure a **data harbour**. A harbour implements a set of protocols that allows it to interact with other harbours.

Contracts: Are a set of rules that are shared amongst participating harbours which describe how objects (data, compute) can be traded between harbours and who can process data. In its simplest form is a 7-tuple which binds a user, data object, compute container, contract, consortium, harbour, and expiry date.

An application: Is a distributed pipeline which can make use of several contracts. The combination of application and contract defines the archetype of the computation i.e. how data and compute are moved to effect computation.

Auditor: A trusted entity that collects audit trails for use in litigation of policy violations.

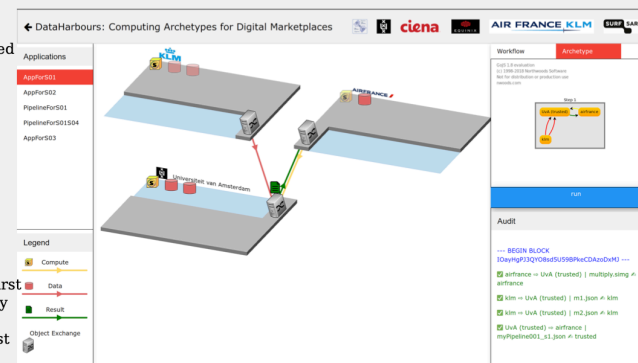


In action

Federated computing on 3 distributed data harbours. Here we illustrate one archetype where KLM and Airfrance do not trust each other and employ a trusted 3rd party to send the data and compute for processing.

For the scenario to succeed the different harbours need to effect several transactions which are governed by contractual rules.

The transaction **protocol** involves first identifying both parties are who they say they are through pub/priv key challenges and secondly, that at least a **contract** rule is matched to allow the transaction. Important steps of the transactions are **audit** logged i.e. signed and published to an audit log collector.



Experimental Setup

Data Transfer nodes at UvA, KLM and Equinix

Running Kubernetes with a number of dockers (pods) see below.

```
tim@uva-kube-04:~$ kubectl get pods -o wide
NAME                                READY   STATUS    RESTARTS   AGE   IP              NODE           NOMINATED NODE
be2-deployment-c87646848-wt815     1/1    Running   0          77m   192.168.5.39   eqx-kube-03    <none>
mq1                                  1/1    Running   0          13d   192.168.1.2    uva-kube-02    <none>
oex.airfrance                       1/1    Running   0          10d   192.168.4.16   eqx-kube-02    <none>
oex.klm                             1/1    Running   0          10d   192.168.3.5    eqx-kube-01    <none>
oex.trusted                         1/1    Running   0          13d   192.168.1.3    uva-kube-02    <none>
planner1                            1/1    Running   0          10d   192.168.4.15   eqx-kube-02    <none>
```

- be2 is the backend for the website.
 - It serves the static pages and passes new information and input to the planner.
- mq1 is the message queue that each oex writes logs to.
- oex.* each is one zone or 'object exchange server'.
 - It is responsible for handling requests from others and sending requests to other parties.
 - It should do so in accordance with a preselected contract/archetype.
- planner1 handles requests from be2 when selecting an archetype (and passes it on to the oex's) and when an application or pipeline is started (and passes it on to the oex).

DataHarbours: Computing archetypes for digital marketplaces

Reggie Cushing, Tim van Zalingen, Joseph Hill, Lu Zhang, Paola Grosso,
Cees de Laat, Leon Gommans, Vijaay Doraiswamy, Purvish Purohit,
Kaladhar Voruganti, Craig Waldrop, Rodney Wilson, Marc Lyonnais

Q&A

- More information:
 - <http://dl4ld.nl>
 - <http://dl4ld.net>
 - <http://sc.delaaat.net/>

