

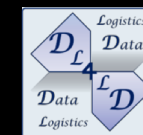
Data Protection Digital Marketplaces Using Novel Infrastructure Models.!

The Global Big Data Hub infrastructure inspired by PRP

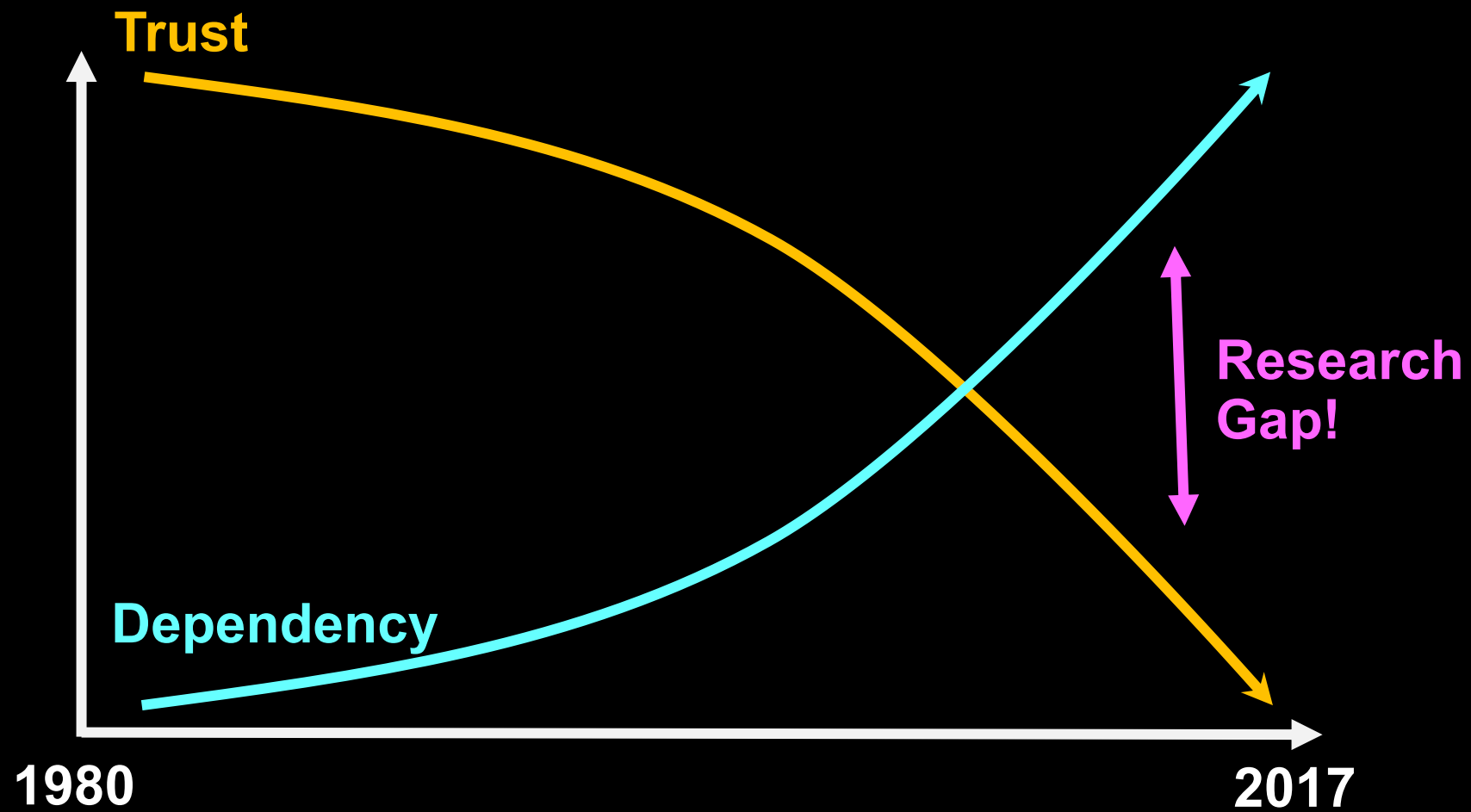
Cees de Laat

Systems & Network Laboratory
University of Amsterdam

APC2018



Fading Trust in Internet



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?

Big Data Sharing use cases placed in airline context



Global Scale



Aircraft Component Health Monitoring (Big) Data
NWO **CIMPLO** project
4.5 FTE

National Scale



Cargo Logistics Data
(C1) DaL4LoD
(C2) **Secure scalable policy-enforced distributed data Processing**
(using blockchain)



Cybersecurity Big Data
NWO COMMIT/
SARNET project
3.5 FTE

City / regional Scale

Campus / Enterprise Scale

NLIP iShare project

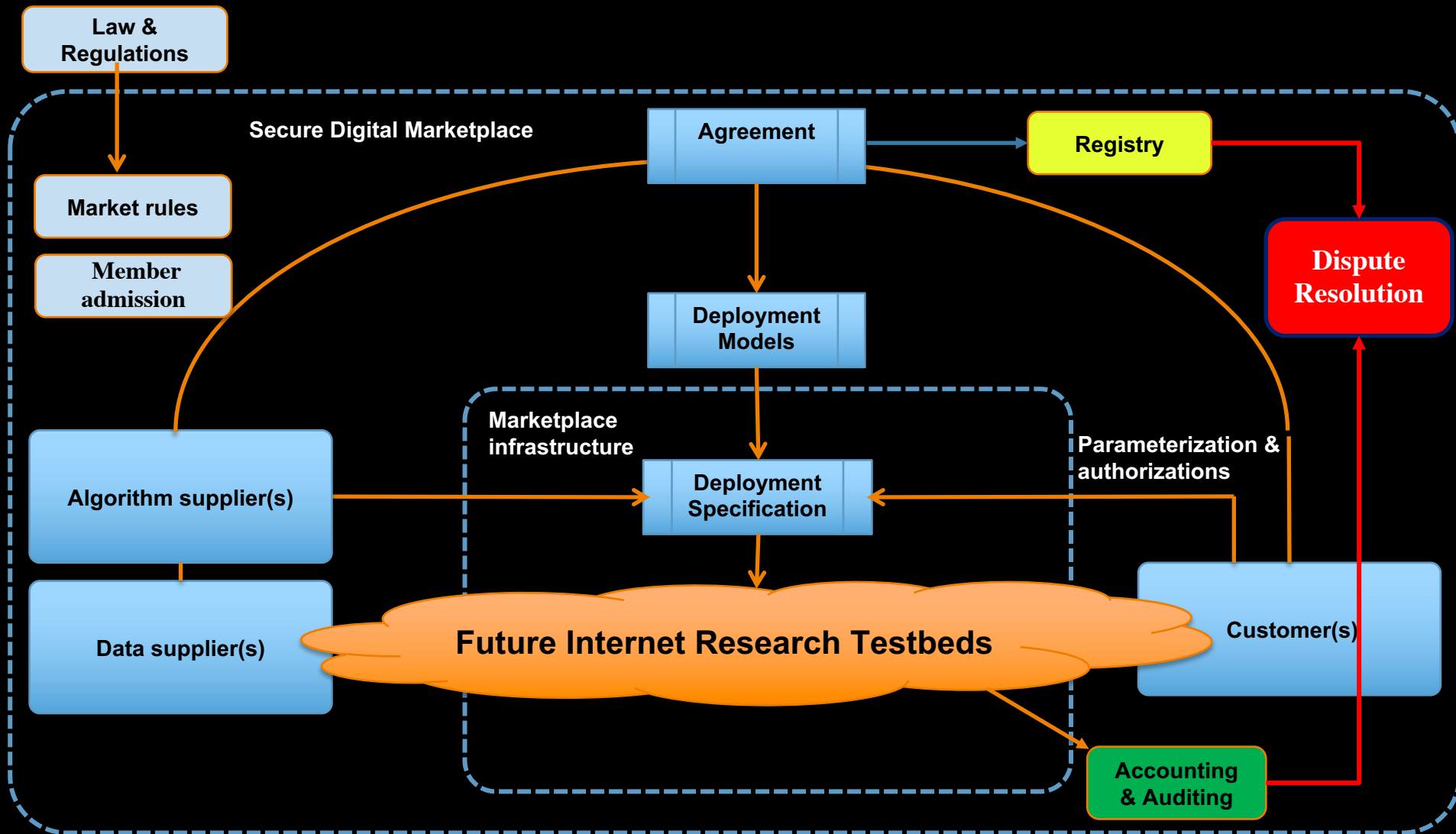


Approach

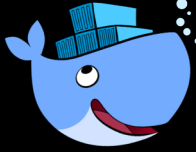
- Strategic:
 - Translate legislation into machine readable policy
 - Define data use policy
 - Trust evaluation models & metrics
- Tactical:
 - Map app given rules & policy & data and resources
 - Bring computing and data to (un)trusted third party
 - Resilience
- Operational:
 - TPM & Encryption schemes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



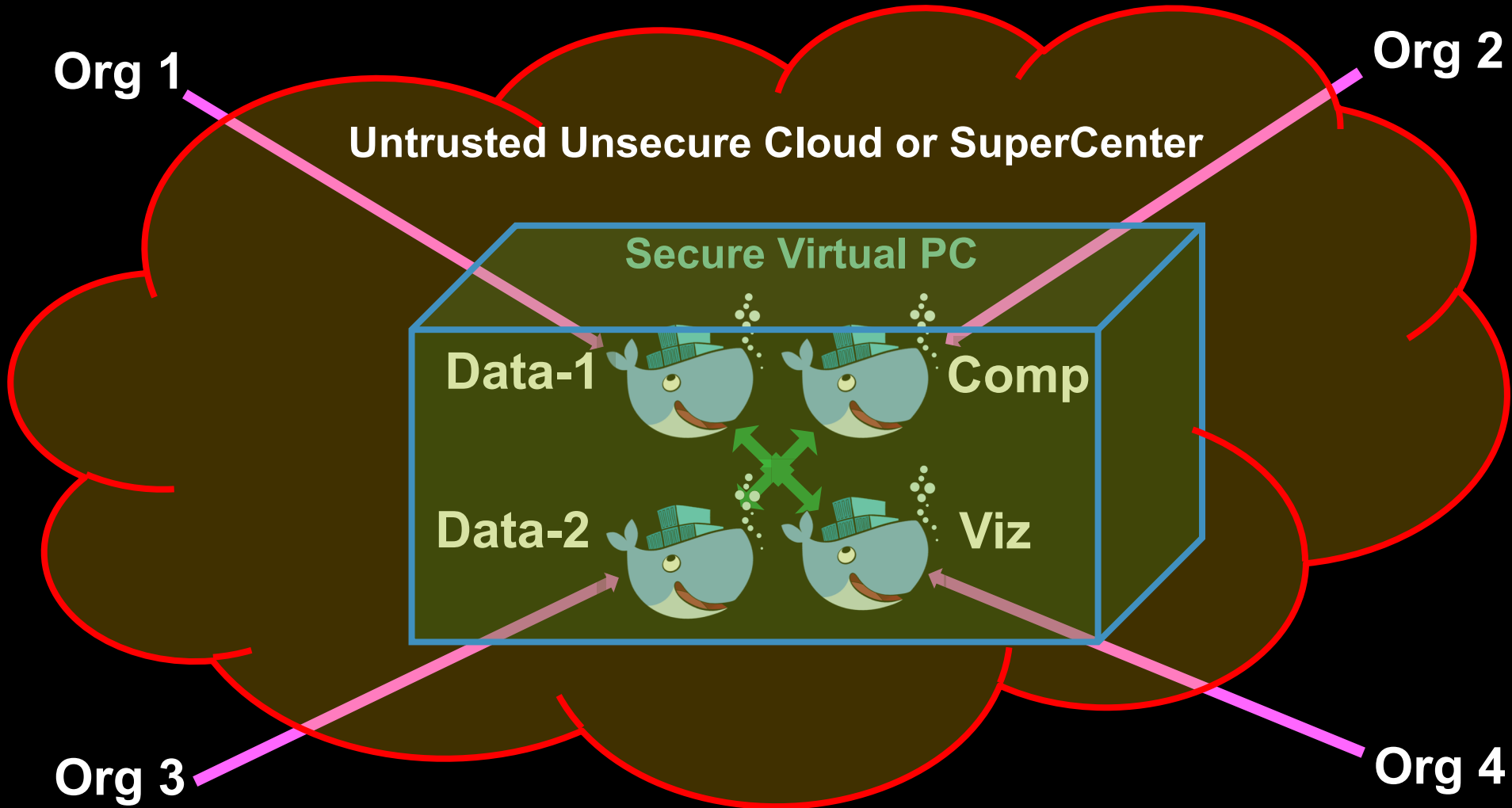
Secure Digital Market Place Research



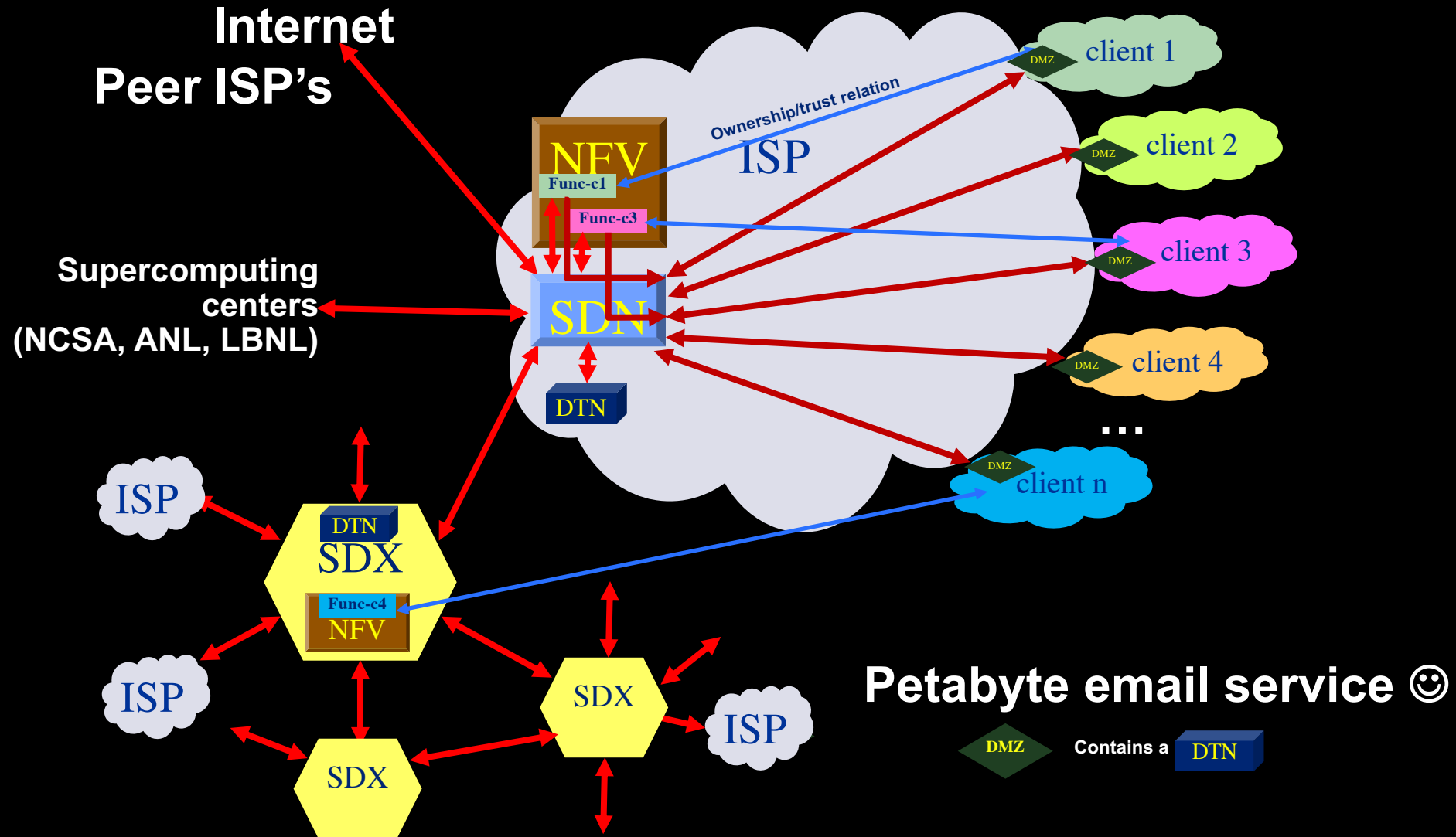
Secure Policy Enforced Data Processing



- 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)



Networks of ScienceDMZ's & SDX's



SC16 Demo

DockerMon

Sending docker containers with search

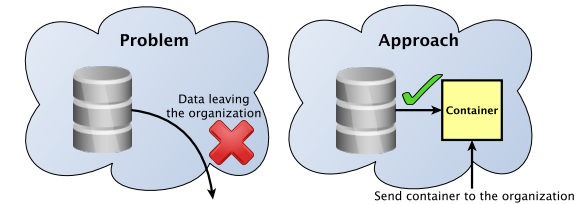
algorithms to databases all over the world.

<http://sc.delaat.net/sc16/index.html#5>

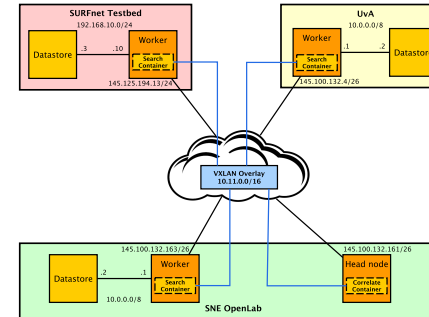
Container-based remote data processing

Problem Description

- Scientific datasets are usually made publicly available
...but data cannot always leave the organization premises
- On-site data processing can be challenging because of incompatibility of systems or lack of manpower
- Can a container-based system perform remote on-site data processing efficiently?
- What are the networking issues to solve?



Underlay and Overlay



Main features:

- Networked containers
- VXLAN overlay
- Containers that perform data retrieval and computation
- Containers built on-demand
- On-site data processing
- Distributed data source
- Multiple sites with datasets

The Game

Our SC16 demo is a gamification of the remote dataset processing architecture.

How many different animal species can you find? You have a fixed budget and each function and processing will cost you money!

In our game you will:

- Select a correlate function to combine the results of the different sites.
- Pick different search functions, represented as tools, to find animals in the remote datasets.
- Build containers with the search and correlate functions.
- Execute the containers on the sites of your choice.

Will you have the best score?



Data Hub System Applicability

Industry

- Cross Cutting Field lab
- Innovation with SURF



Science

- European Open Science Cloud
- FAIR model
 - Findable – Accessible – Interpretable - Reusable



Society

- Smart Cities & Arena
- Streaming Data Decision Support

