

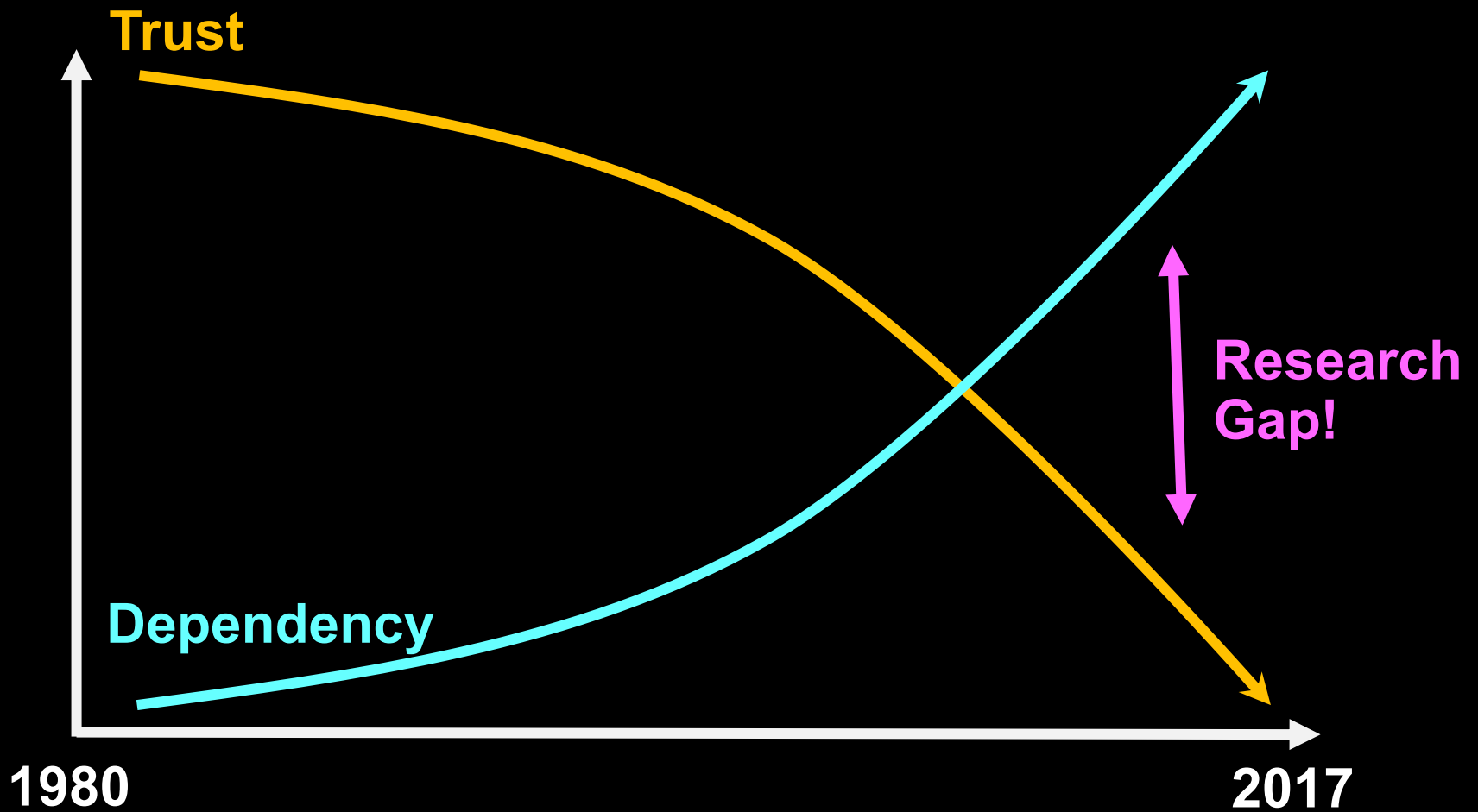
Secure Cyber Infrastructure for Valuable Big Data Processing!

Cees de Laat

Systems & Network Laboratory
University of Amsterdam



Fading Trust in Internet

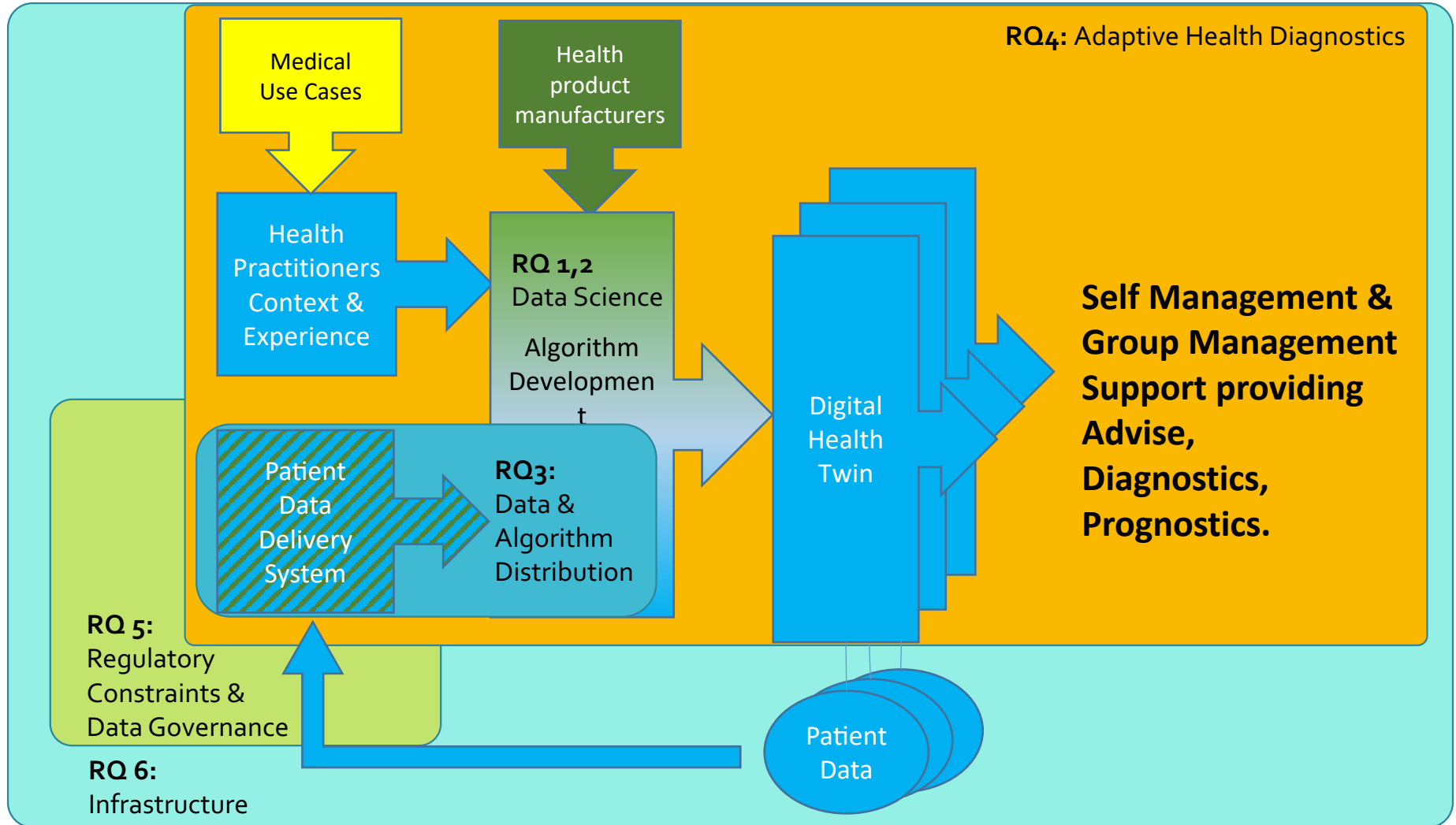


Main problem statement

- 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?

Health use case

Enabling Personal Interventions



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



iSHARE
powered by NLIP



SAE Use Case envisaged research collaboration

Funding Agency



Big Data Hub / Spoke or Industry initiative funding



International Networking



Regional / National Networking



Local University



Aircraft MRO, OEM & Operators



Industry Standards Body



SAE AeroSpace Group
HM-1 working group
Use Case on aircraft sensor Big Data

ECONOMY

Managing Our Hub Economy

by [Marco Iansiti](#) and [Karim R. Lakhani](#)

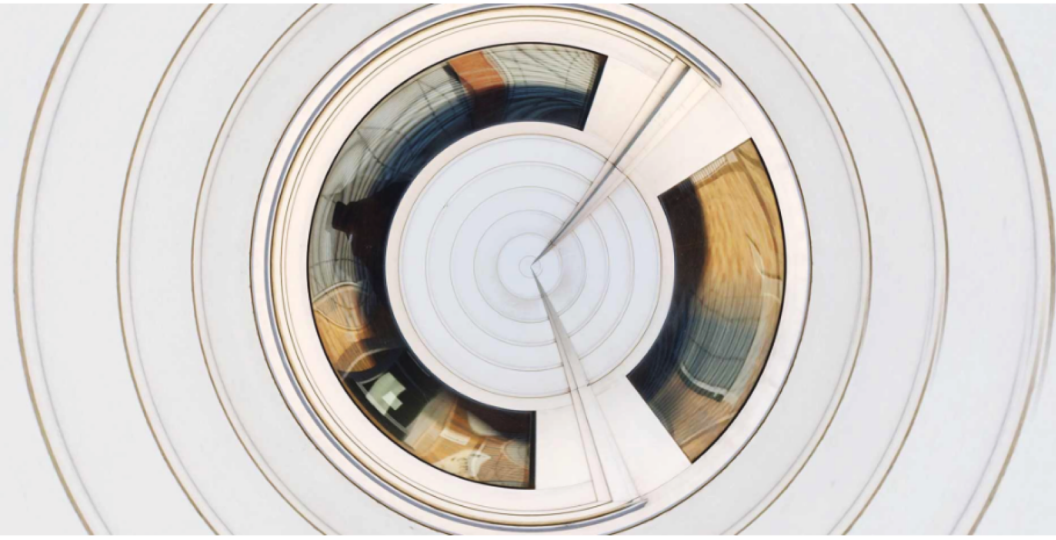
FROM THE SEPTEMBER-OCTOBER 2017 ISSUE

WHAT TO READ NEXT



[The IT Transformation Health Care Needs](#)

SUMMARY SAVE SHARE COMMENT 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

data Sharing approach: Combine 2&3

MANAGED BY AN INDEPENDENT INDUSTRY MEMBERSHIP ORGANIZATION

Case studies

Approaches to B2B data sharing



Five different approaches to B2B data sharing

1 DATA MONETISATION



2 DATA MARKETPLACES



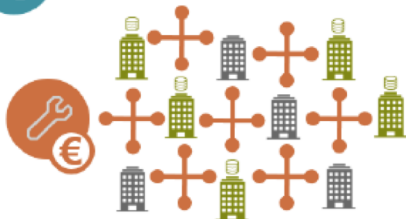
3 INDUSTRIAL DATA PLATFORMS



INDUSTRIAL DATA PLATFORMS

- ✓ Strategic and collaborative partnerships
- ✓ Mutual benefits for all parties
- ✓ Data shared (for free) in a closed, exclusive and secure environment
- ✓ Develop new or improved products and/or services
- ✓ Enhance internal performance

4 TECHNICAL ENABLERS



5 OPEN DATA



Independent Marketplace data platform governed by industry membership organization

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!)

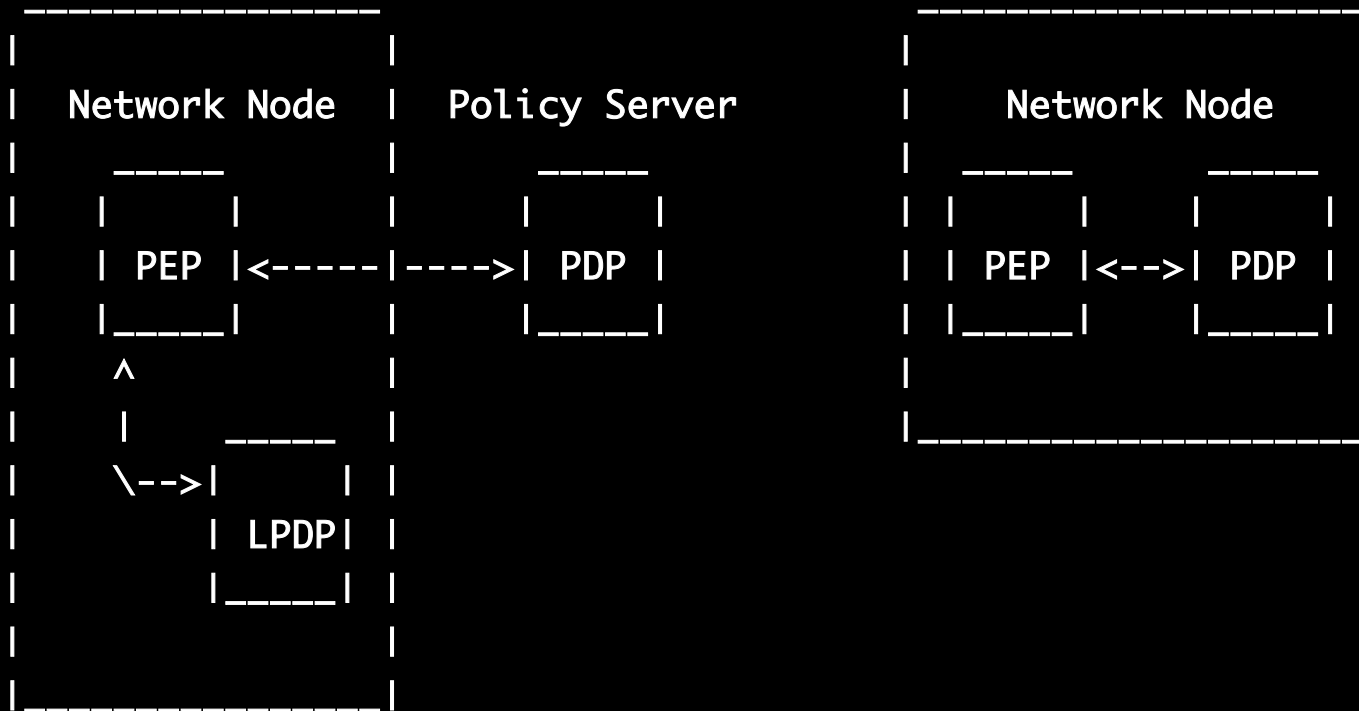


Line of research

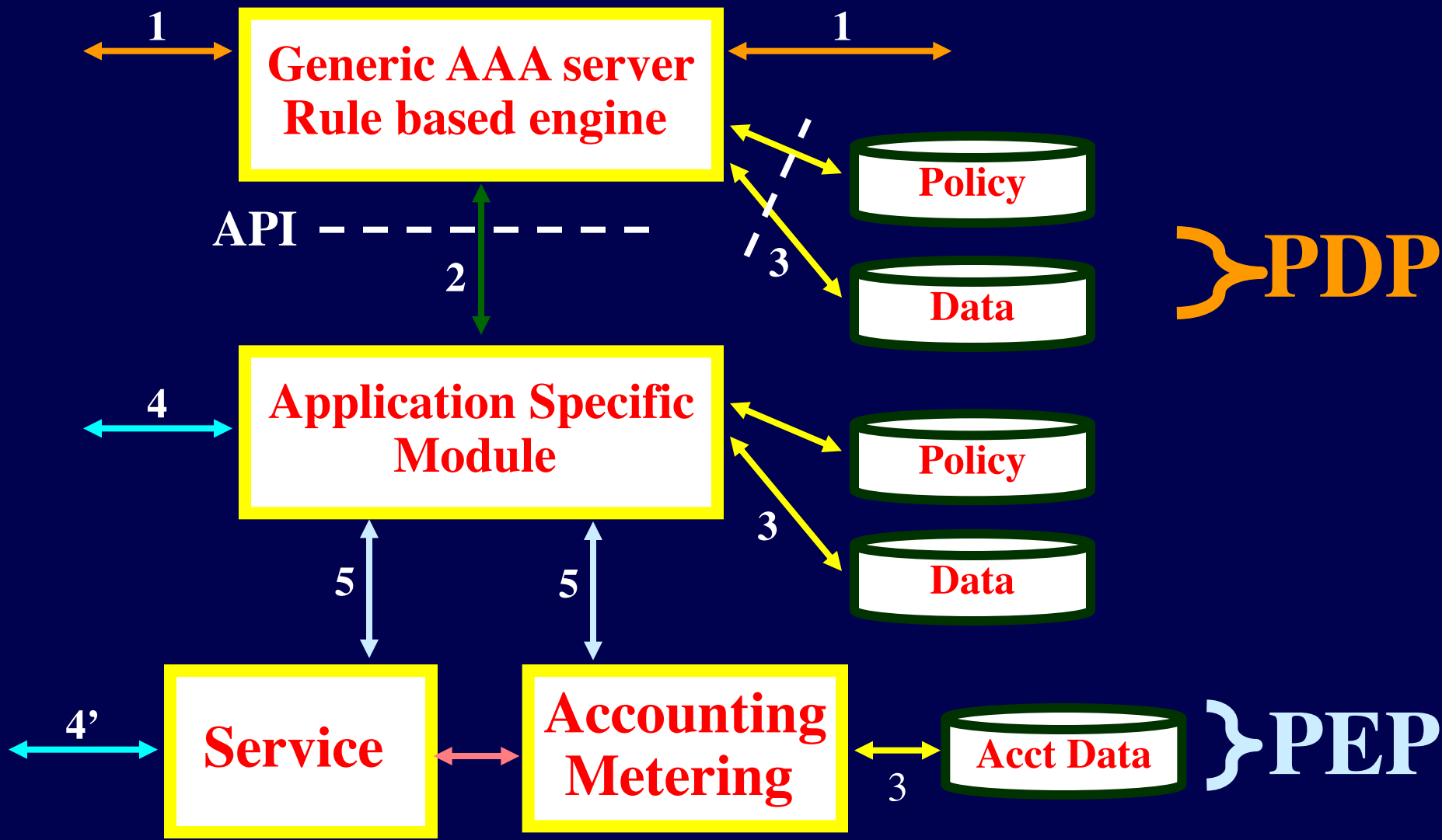
- 1997: Need for authorization framework for combination of resources across domains
- 1998: AAA-ARCHitecture research in IRTF
- 2000: RFC 2903-2906, 3334
- 2005: open versus not so open exchanges
- 2006: start of trust research (also in rfc 2904)
- 2012: I2-spring session presenting line of research
- 2014: PhD defense of research plus publication
- 2015: SARNET organizing trust across domains
- 2018: DL4LD and EPI Big Data Markets

IETF: Common Open Policy Service (COPS)

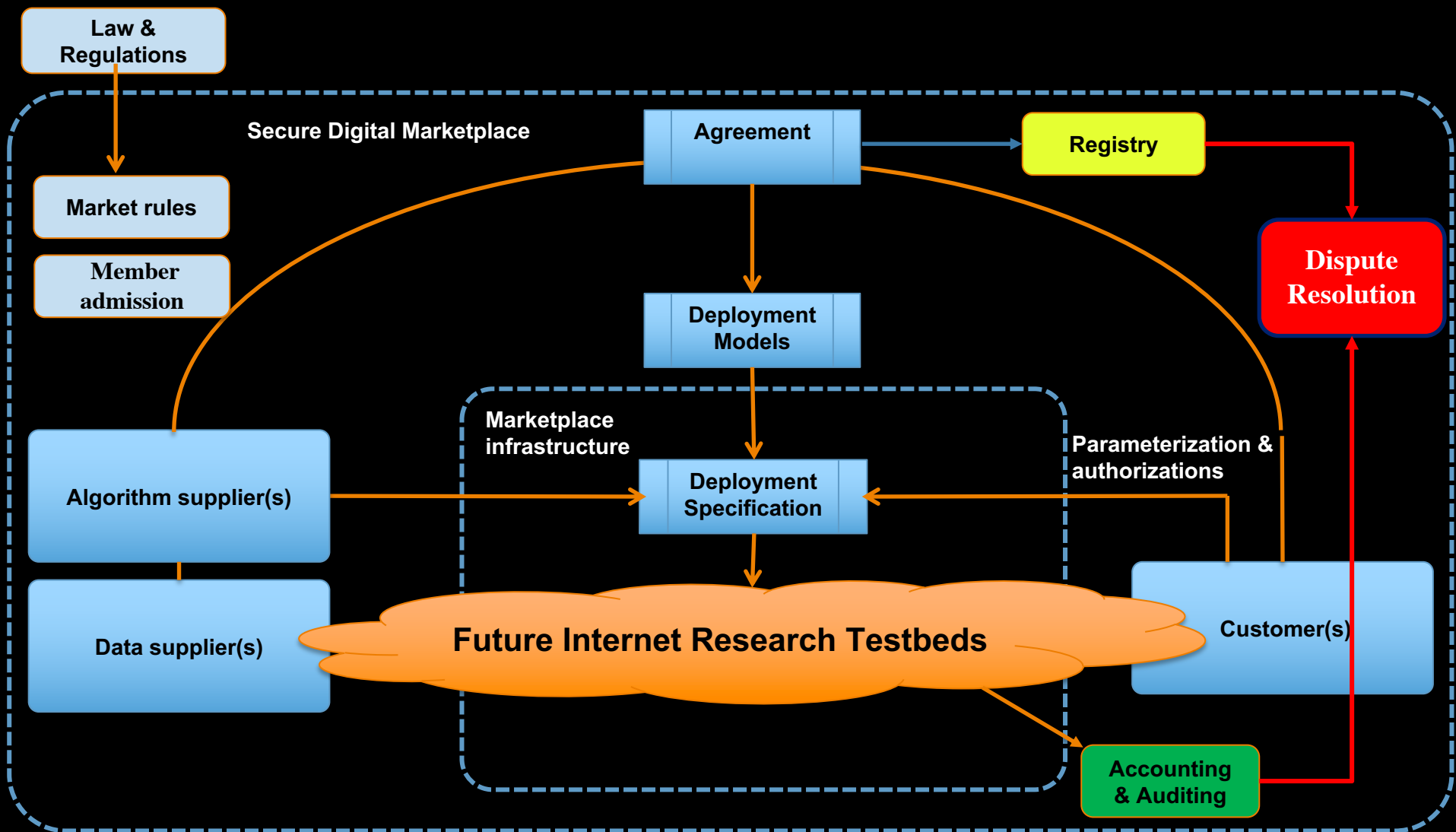
- Rfc 2748, 2753, 4261



IETF/IRTF: AAA Architecture research



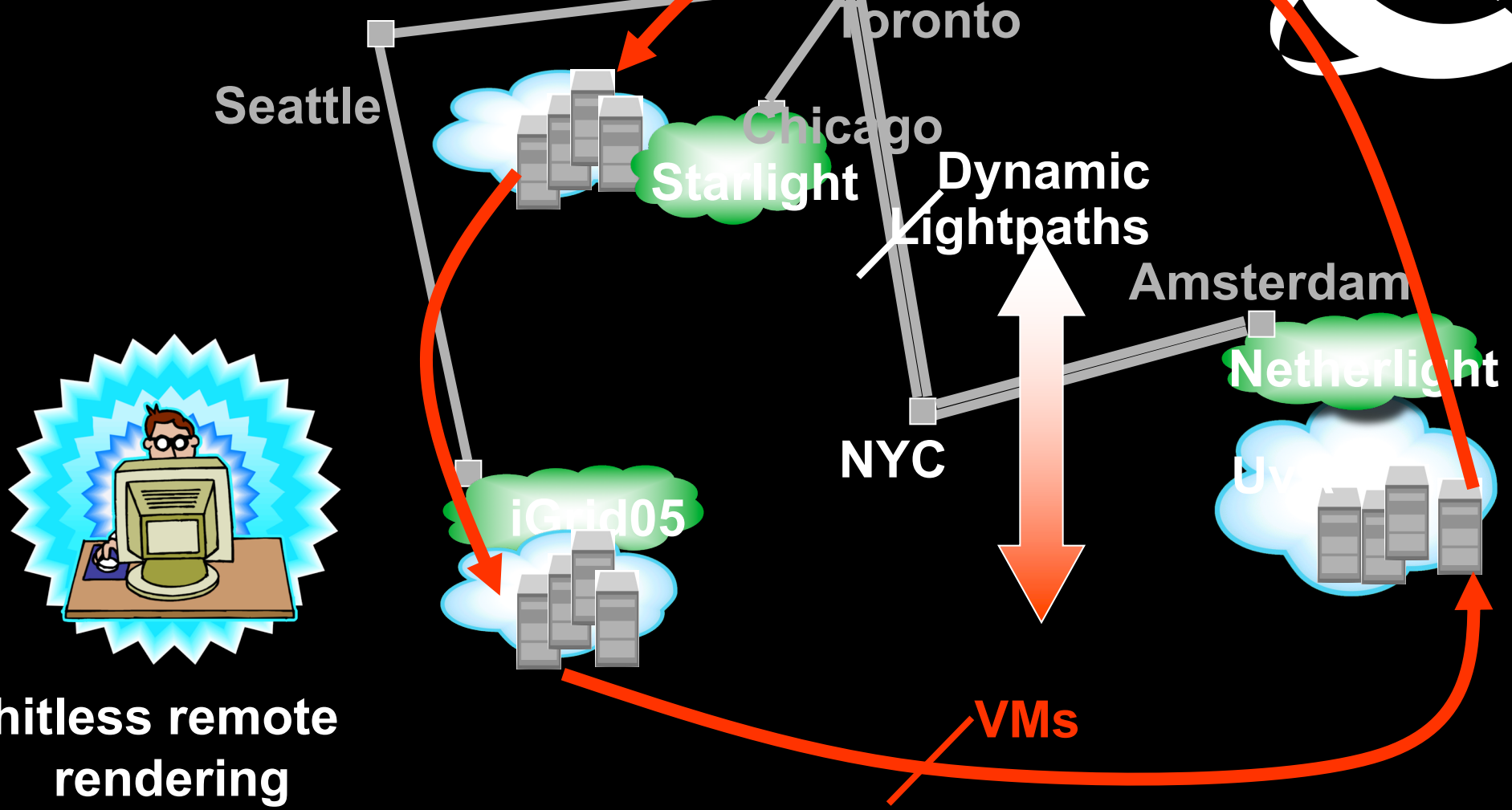
Secure Digital Market Place Research



Data Processing models

- Bring data to computing
- Bring computing to data
- Bring computing and data to (un)trusted third party
- A mix of all of the above
- Block chain to record what happened
- Block chain for data integrity
- Bring the owner of Data in control!
- Data owner policy + enforcement technology

The VM Turntable Demonstrator



hitless remote rendering

The VMs that are live-migrated run an iterative search-refine-search workflow against data stored in different databases at the various locations. A user in San Diego gets hitless rendering of search progress as VMs spin around

Experiment outcomes

Note, this was in 2005 at SC and igrid2005!



We have demonstrated seamless, live migration of VMs over WAN

For this, we have realized a network service that

- Exhibits predictable behavior; tracks endpoints

- Flex bandwidth upon request by credited applications

- Doesn't require peak provisioning of network resources

Pipelining bounds the downtime in spite of high RTTs

- San Diego – Amsterdam, 1GE, RTT = 200 msec, downtime \leq 1 sec

- Back to back, 1GE, RTT = 0.2-0.5 msec, downtime = \sim 0.2 sec*

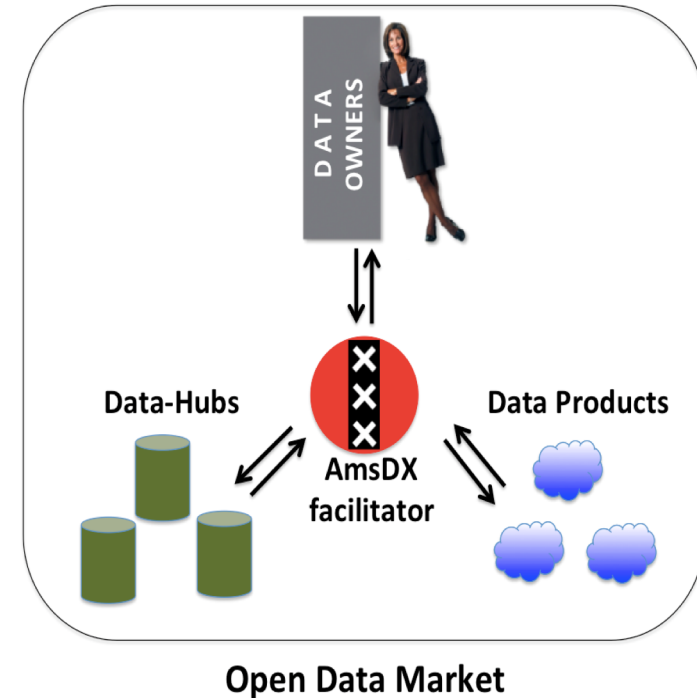
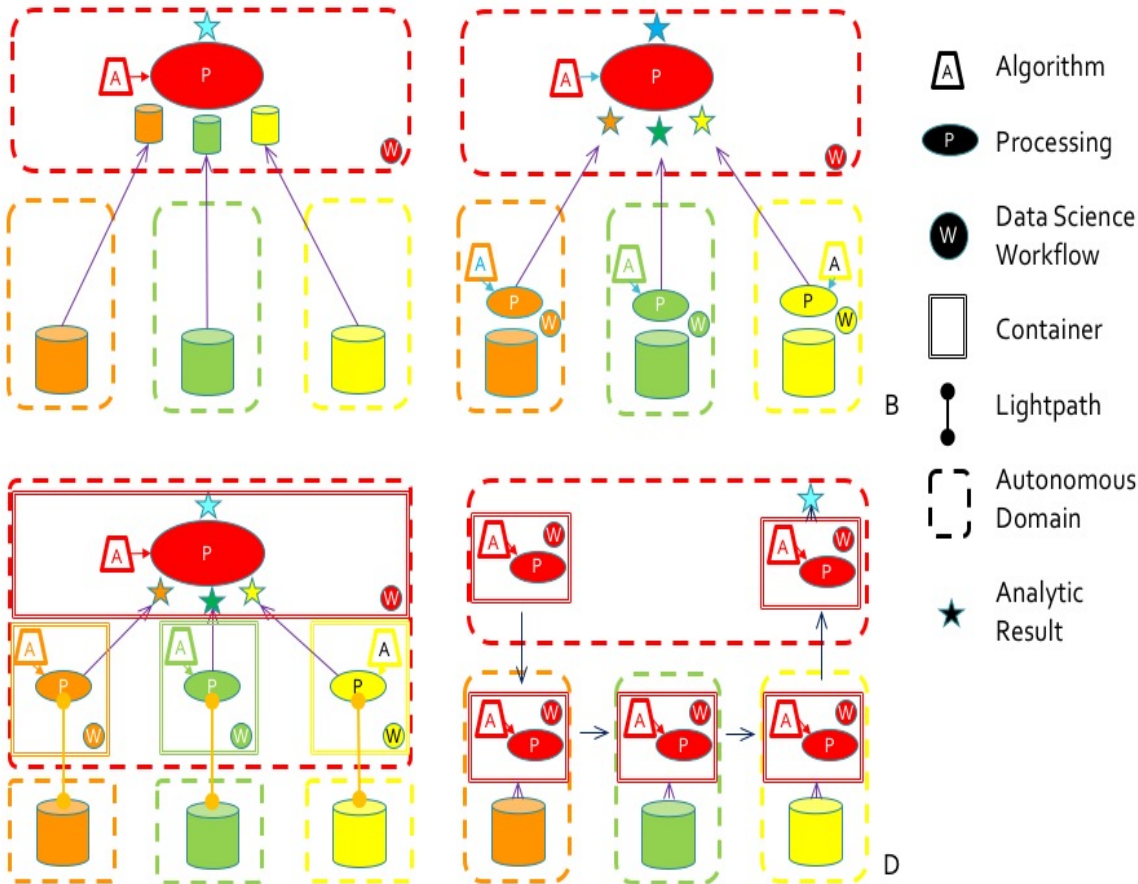
**Clark et al. NSDI 05 paper. Different workloads*

VM + Lightpaths across MAN/WAN are deemed a powerful and general alternative to RPC, GRAM approaches

We believe it's a representative instance of active cpu+data+net orchestration

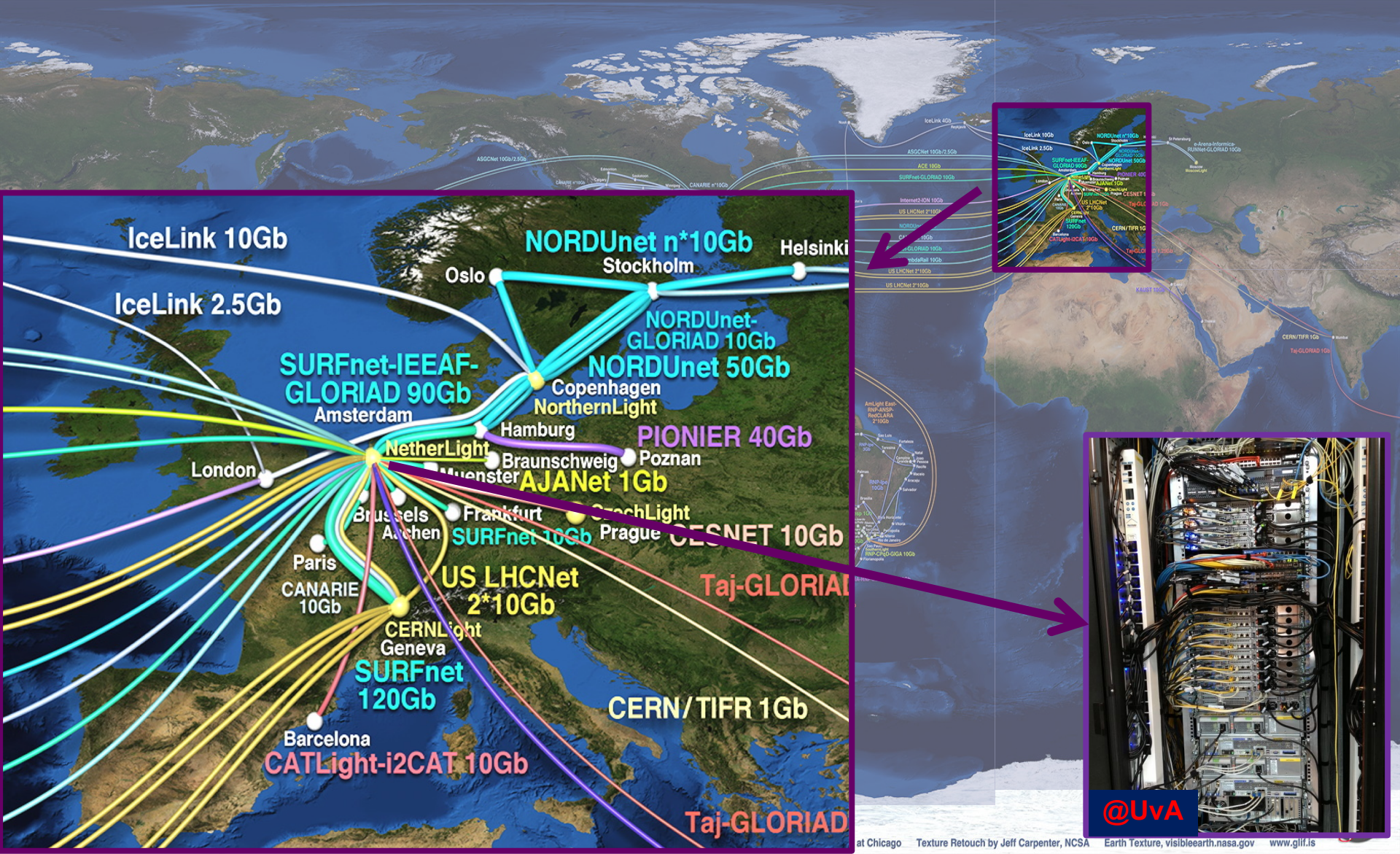
INFRASTRUCTURE PATTERN EXAMPLES

OFFERED BY A DATA EXCHANGE TO MARKETPLACES TO CHOOSE FROM



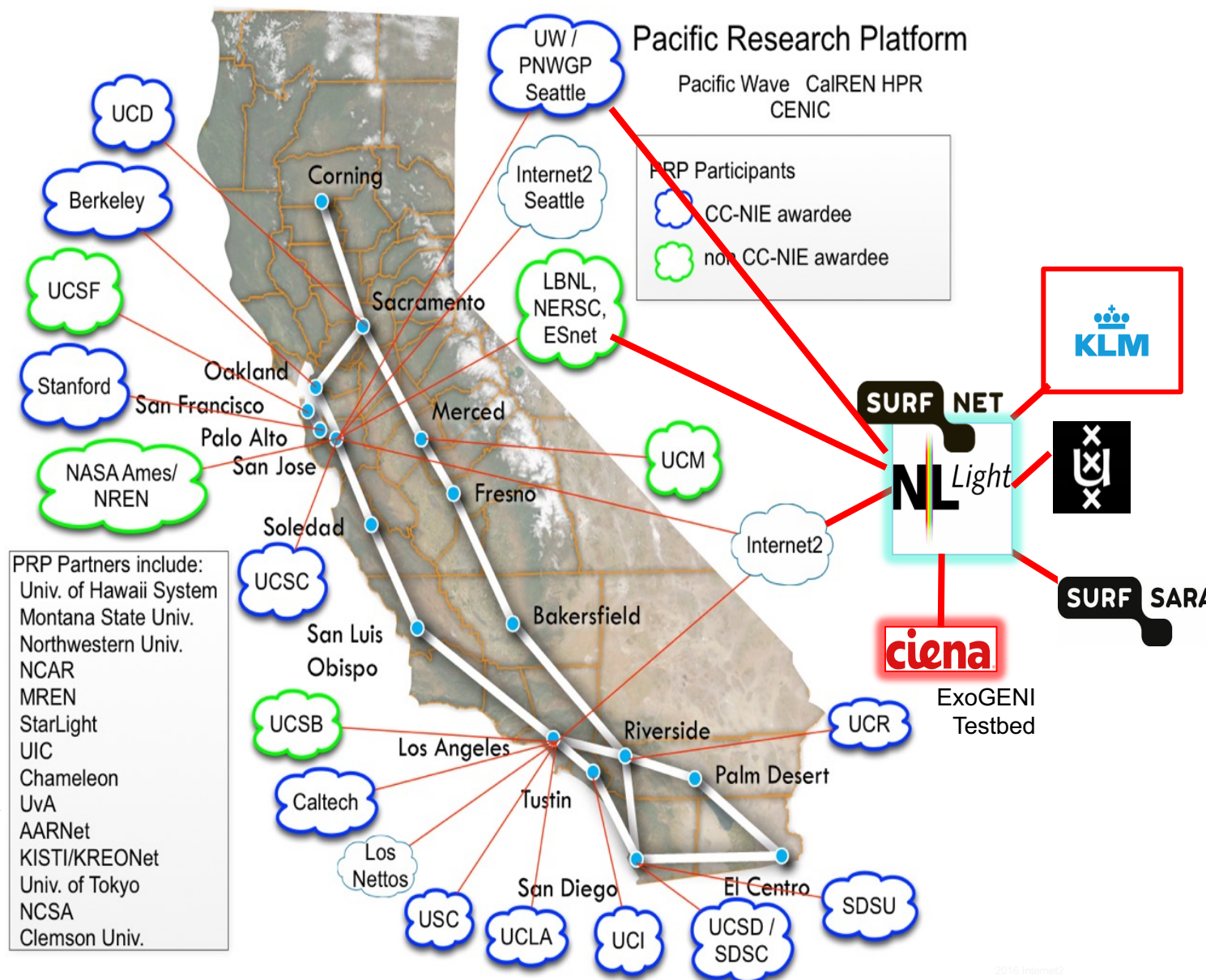
Amsterdam is a major hub in The GLIF

F Dijkstra, J van der Ham, P Grosso, C de Laat, "A path finding implementation for multi-layer networks", Future Generation Computer Systems 25 (2), 142-146.

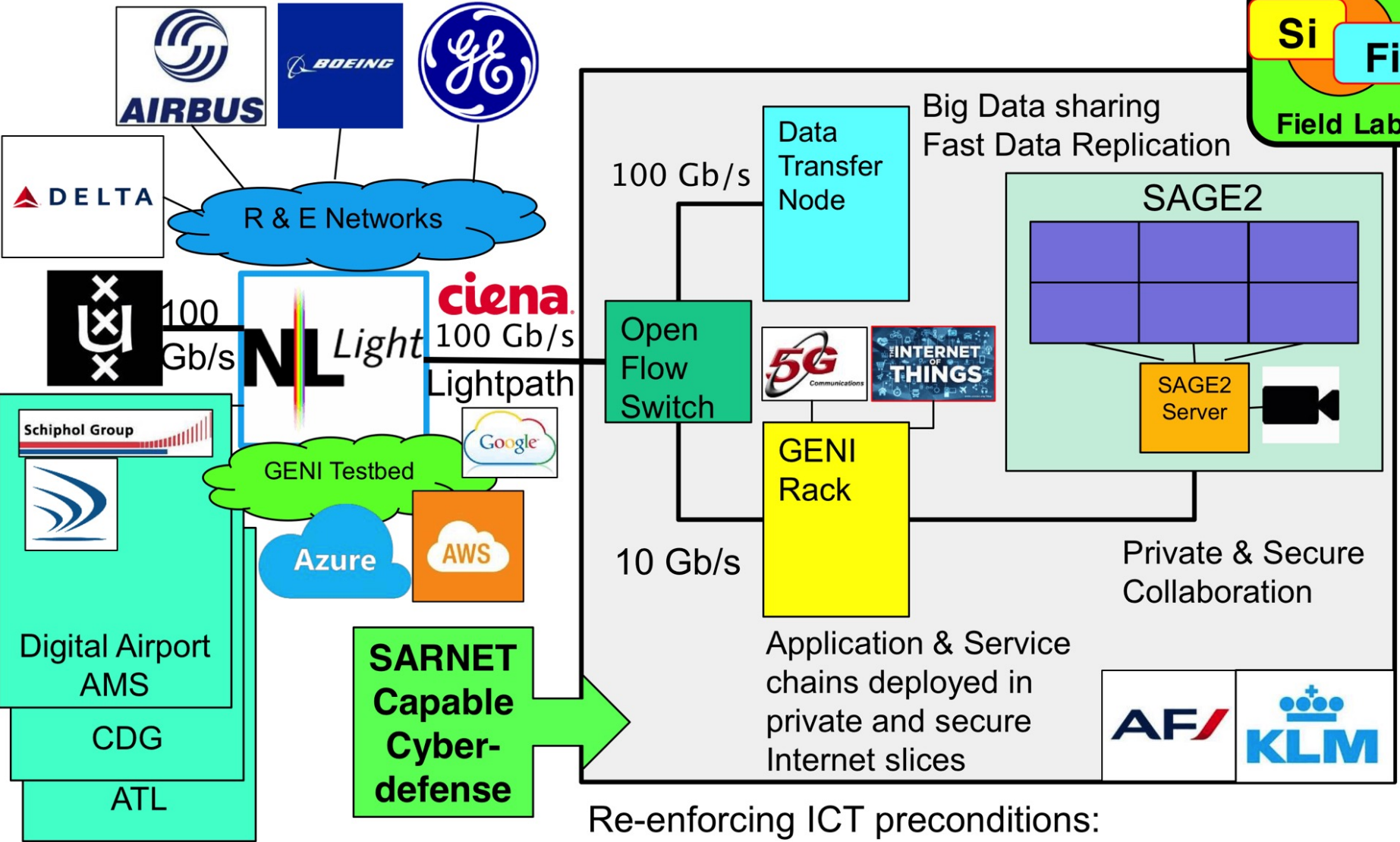
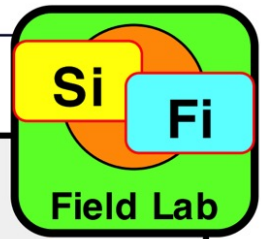


Pacific Research Platform testbed involvement

Research goal:
Explore value of academic network research capabilities that enable innovative ways & models to share big data assets

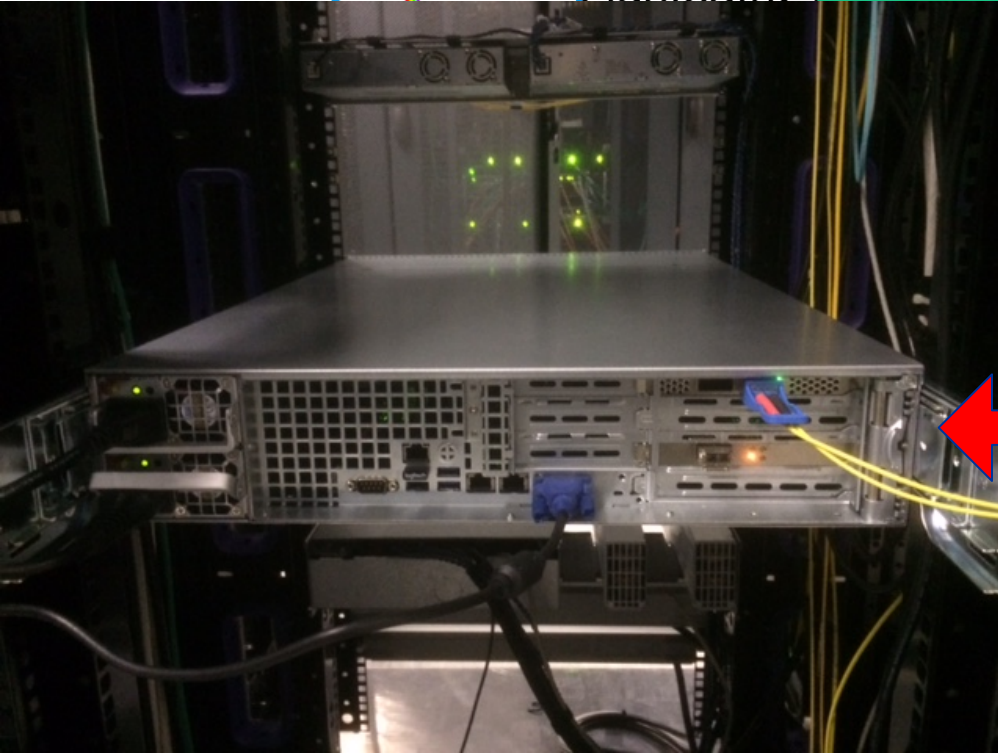
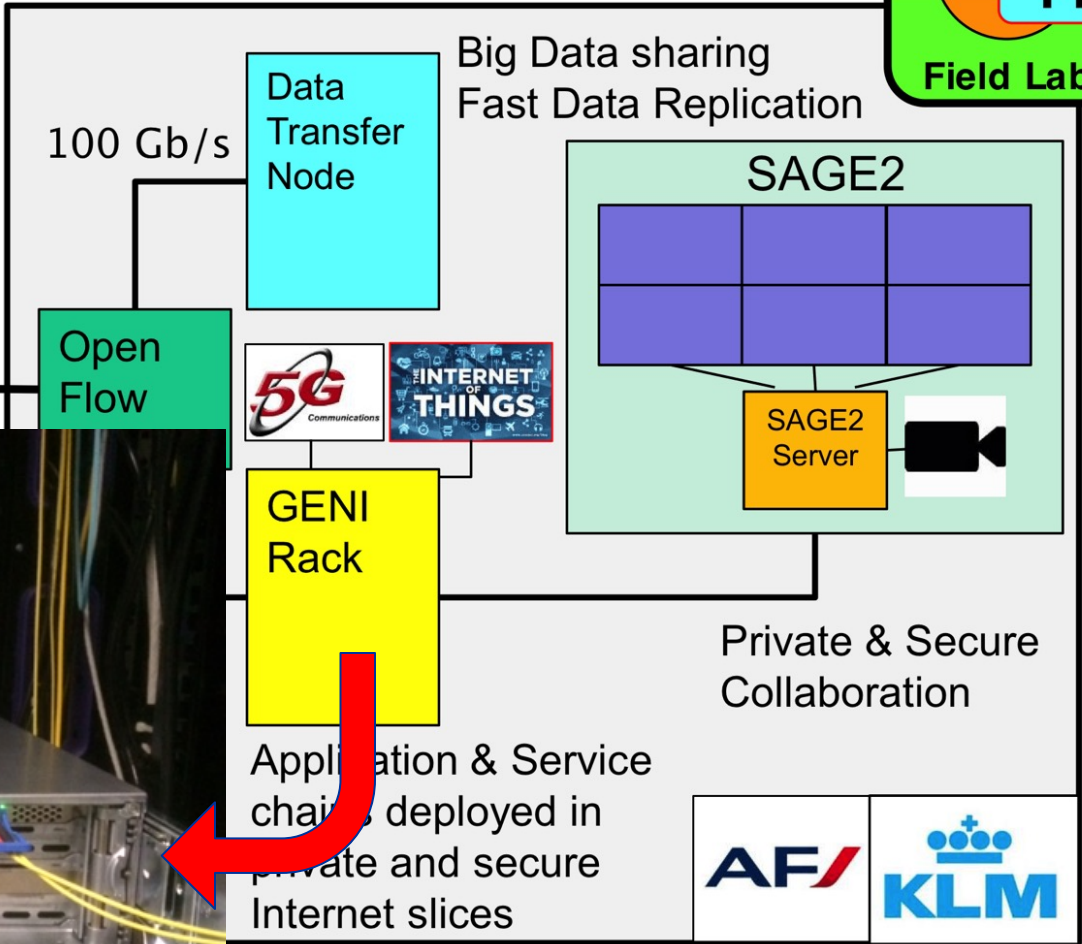
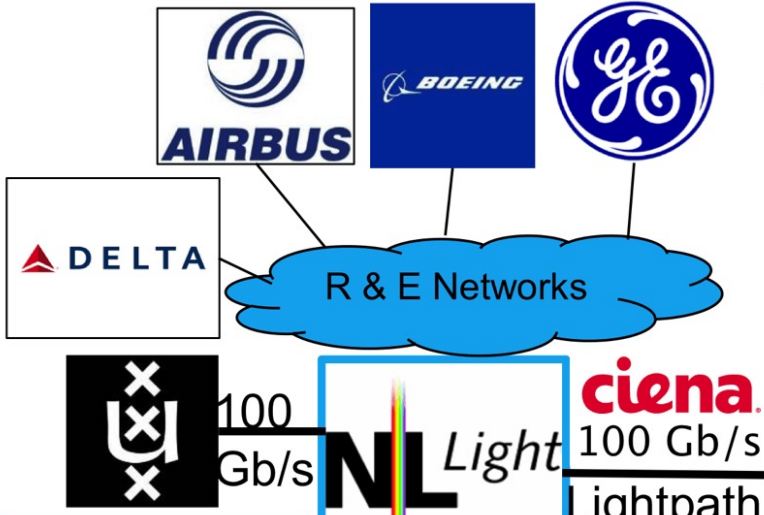
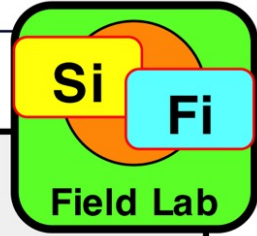


Ambition to put capabilities into fieldlab

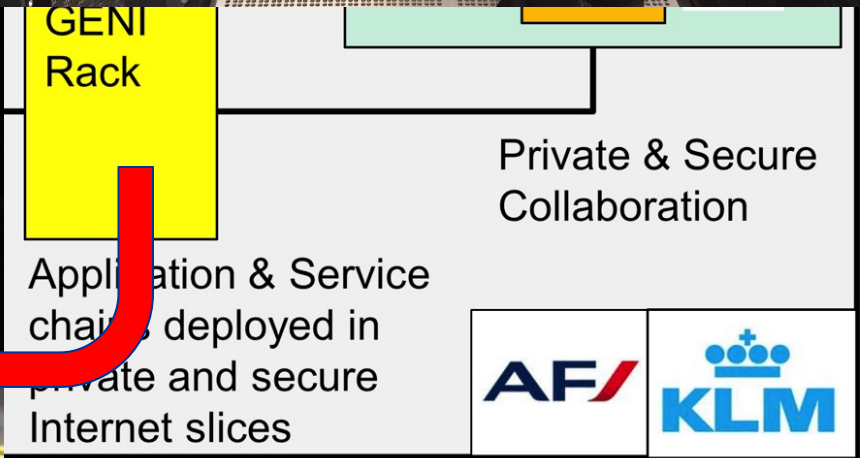


Re-enforcing ICT preconditions:
Each envisaged site has similar elements

Ambition to put capabilities into fieldlab



ing ICT preconditions:
saged site has similar elements

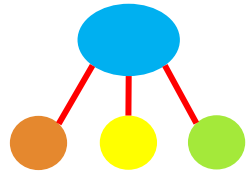


ing ICT preconditions:
saged site has similar elements

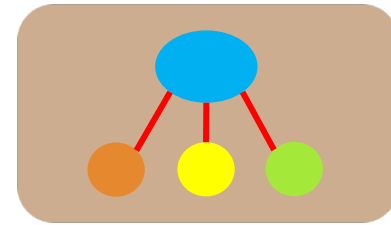
RESEARCH WORKING ALONGSIDE IT INDUSTRY

NETWORK RESEARCH INFRASTRUCTURES

Data Sharing Infrastructure Model
 Research using Future Internet capabilities



COMMERCIAL DATACENTER INFRASTRUCTURE AS NEUTRAL GROUND

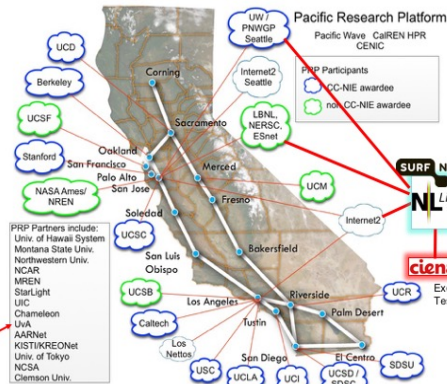


Goal: How to create a Digital Marketplace Ecosystem



prp.ucsd.edu

As foundation of the National Research Platform

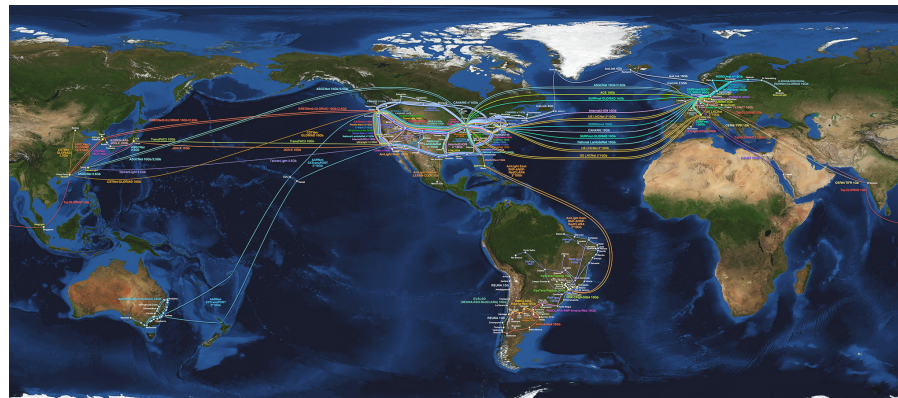


Data Transfer Node at KLM fieldlab with 100 gb/s link to enable SDMP research thanks to UvA, SURFnet and Ciena



AM3 and AM4 Datacenters Science Park Amsterdam

SV10 Datacenter Silicon Valley



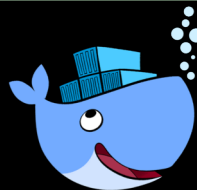
Exchange

Data objects & methods
Data & Algorithms service

Routers - Internet – ISP's - Cloud
IP packet service

Layer 2 exchange service
Ethernet frames

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)

Org 1

Org 2

Untrusted Unsecure Cloud or SuperCenter

Secure Virtual PC

Data-1

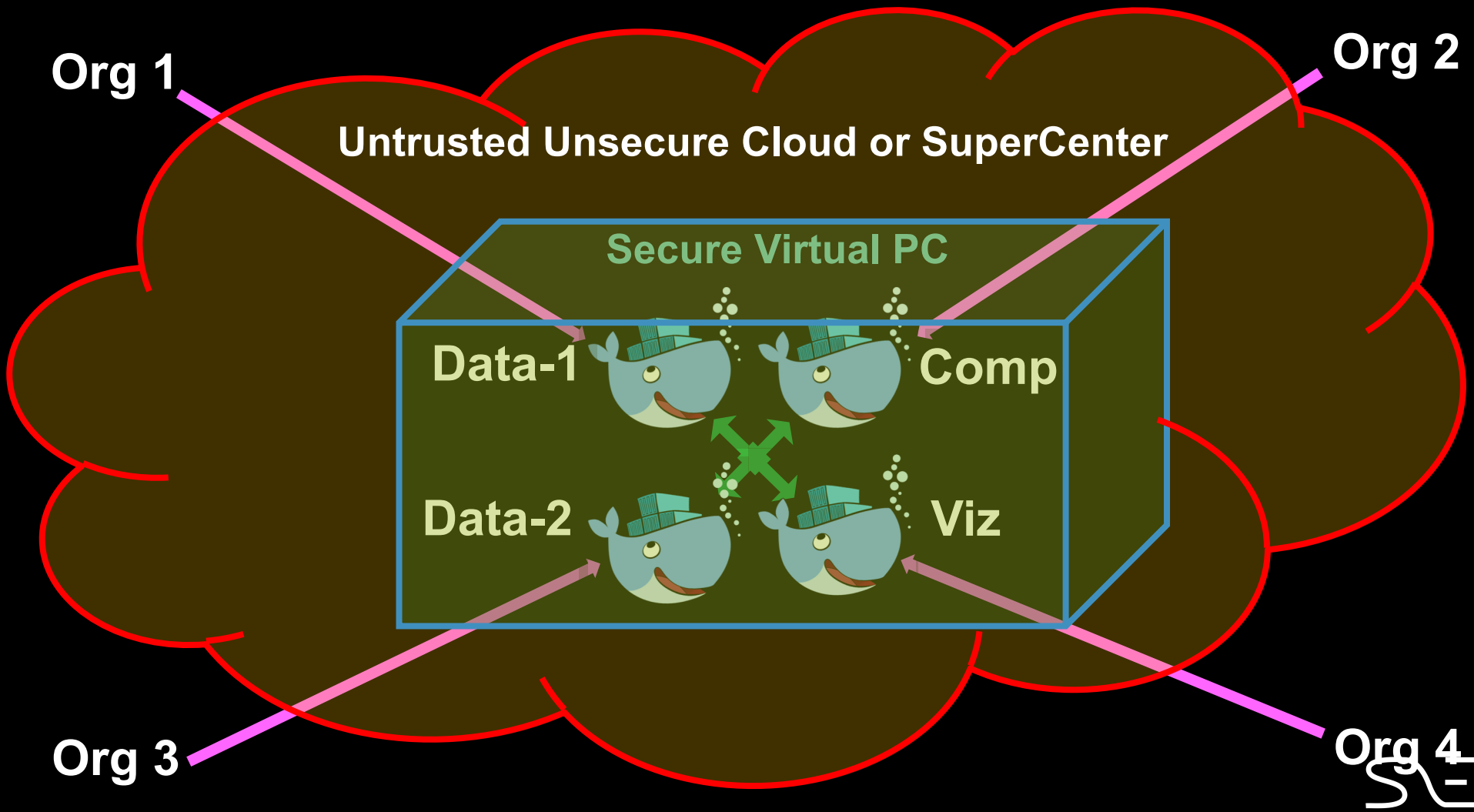
Comp

Data-2

Viz

Org 3

Org 4



SC16 Demo

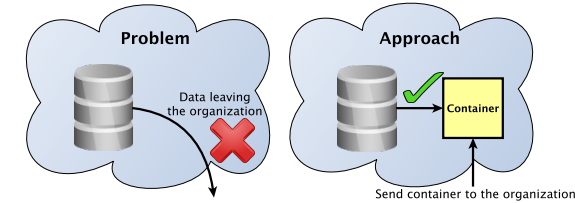
DockerMon Sending docker containers with search algorithms to databases all over the world.

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

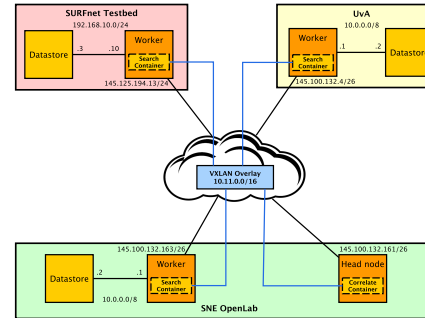
Container-based remote data processing

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.

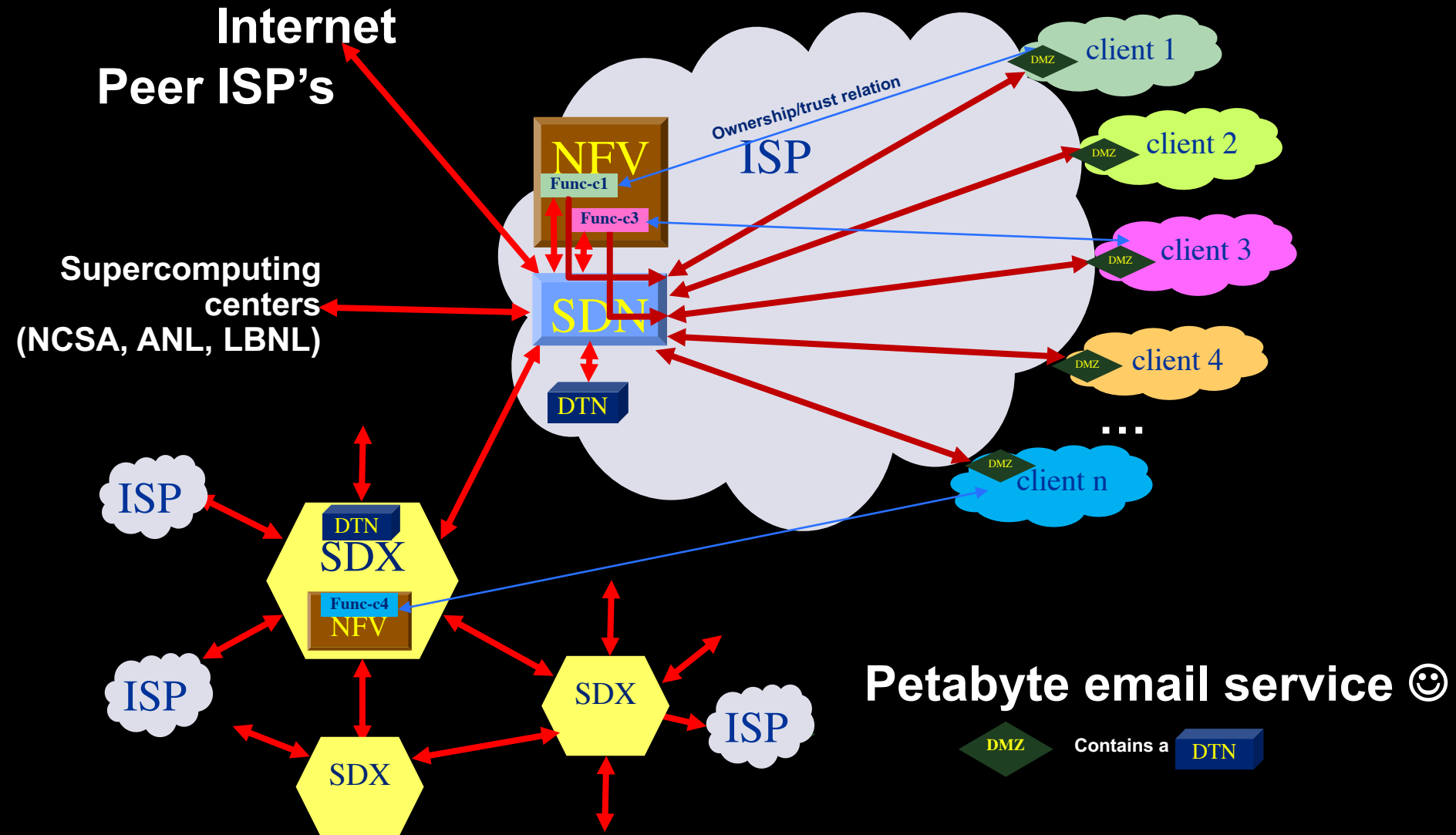
Will you have the best score?



More information:

- <http://byoc.lab.uvalight.net/info>
- <http://sne.science.uva.nl/sne/gigaport3>
- <http://delaat.net/sc>

Networks of ScienceDMZ's & SDX's



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



Q&A



- More information:
 - <http://delaat.net/dl4d>
- Contributions from:
 - Leon Gommans, Tom van Engers, Rob Meijer, Wouter Los and many others!

