

Connecting Data and Algorithms Globally!

The Global Big Data Hub infrastructure inspired by PRP

@ SAE meeting, Palo Alto (CA), Sep 12, 2017

Cees de Laat

System & Network Engineering
University of Amsterdam



Supported by NWO and C2D grants SARNET, DL4LD and NWA.



Science Faculty @ UvA

Informatics Institute



- AMLAB: Machine Learning (Prof. dr. M. Welling)
- FCN: Federated Collaborative Networks (Prof. dr. H. Afsarmanesh)
- ILPS: Information and Language Processing Systems (Prof. dr. M. de Rijke)
- ISIS: Intelligent Sensory Information Systems (Prof. dr. ir. A.W.M. Smeulders)
- CSL: Computational Science Laboratory (Prof. dr. P.M.A. Sloot)
- SNE: System and Network Engineering (Prof. dr. ir. C.T.A.M. de Laat)
- TCS: Theory of Computer Science (Prof. dr. J.A. Bergstra)



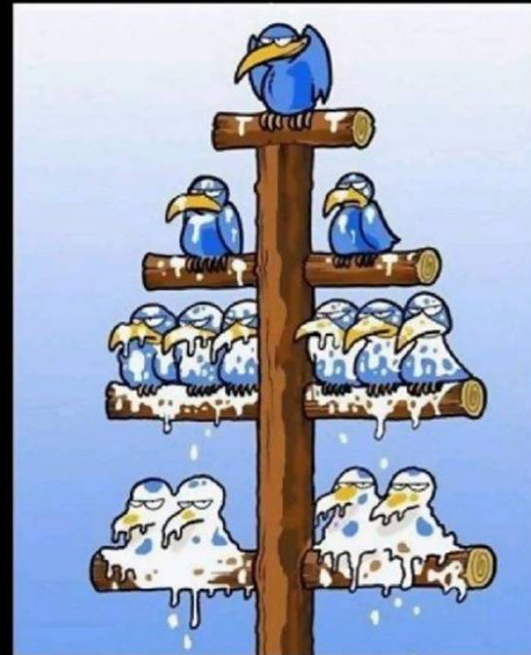
SNE - Staffing

Group leader: prof.dr.ir. C. de Laat

Deputy group leaders: dr. Paola Grosso, dr. Andy Pimentel

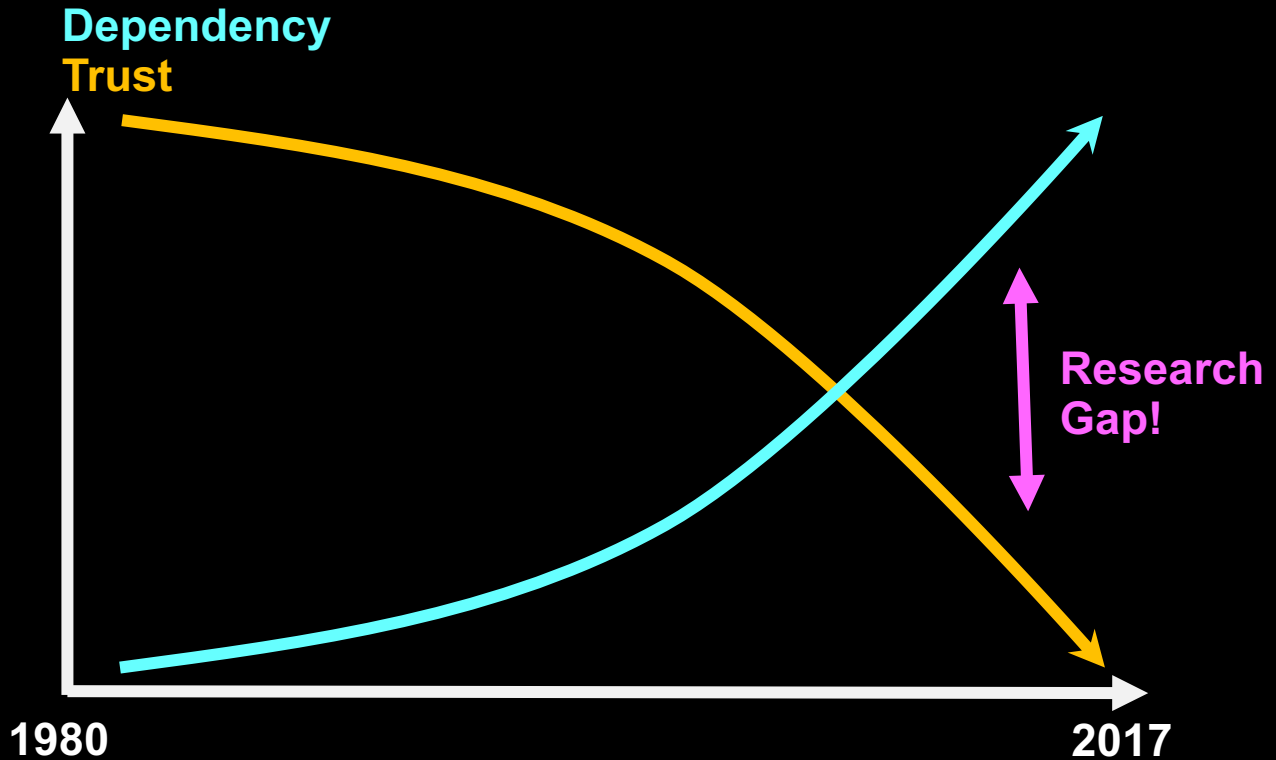
- 1 full prof (me)
- 2 part time professors
- 3 endowed professors
- 2 *senior researchers*
- 2 associate professors
- 3 assistant professors
- ~12 postdoc's
- *About 15 phd students*
- ~10 guests
 - a.o. dr. Leon Gommans
- *Yearly turnover ~ 3 MEuro*

When top level guys look down
they see only shit.



When bottom level guys look up
they see only assholes.

Fading Trust in Internet



Mission

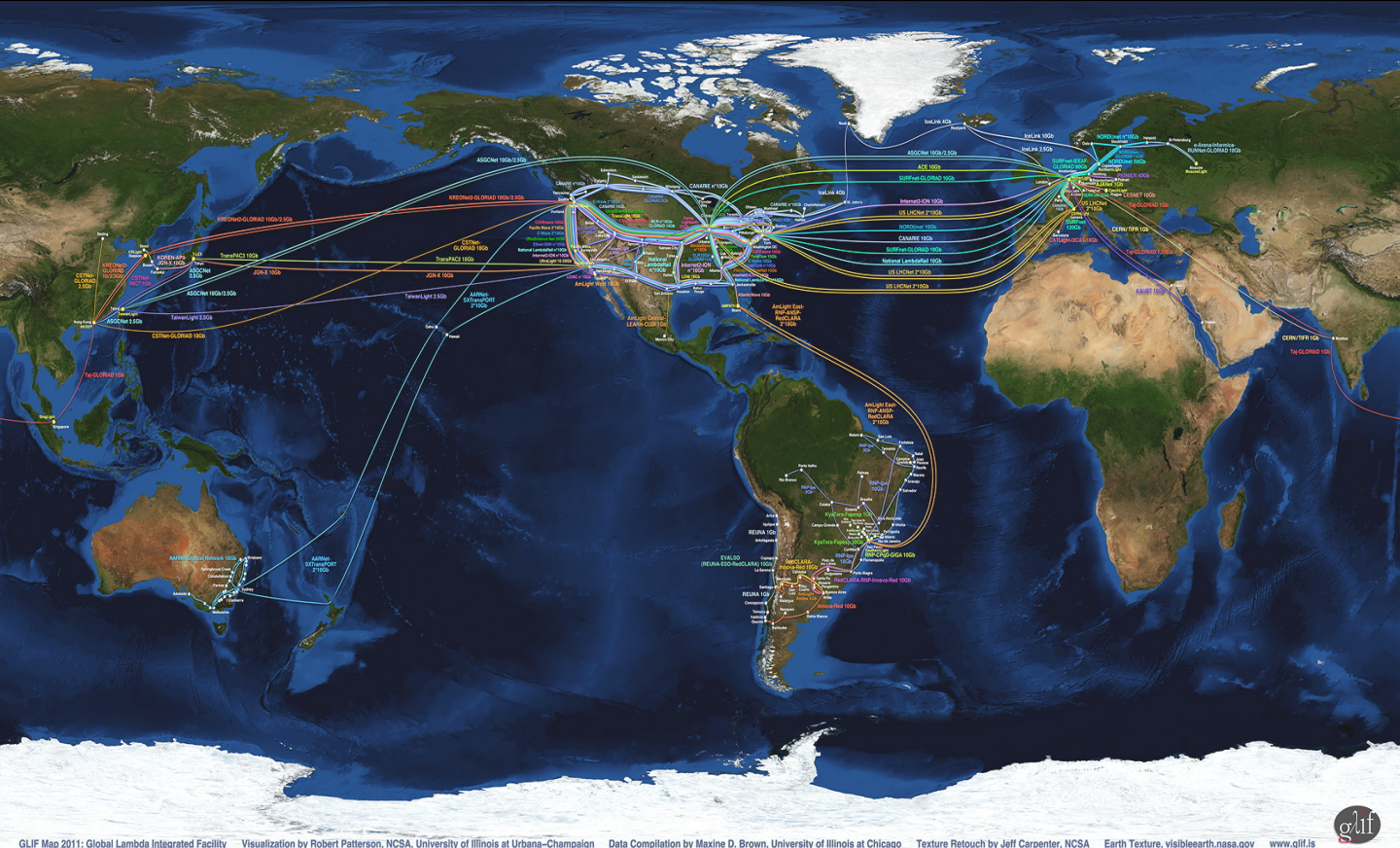
Can we create smart and safe data processing systems that can be tailored to diverse application needs?

- *Capacity*
 - *Bandwidth on demand, QoS, architectures, photonics, performance*
- *Capability*
 - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
 - *Policy, integrity of data in distributed data processing*
- *Sustainability*
 - *Greening infrastructure, awareness*
- *Resilience*
 - *Systems under attack, failures, disasters*



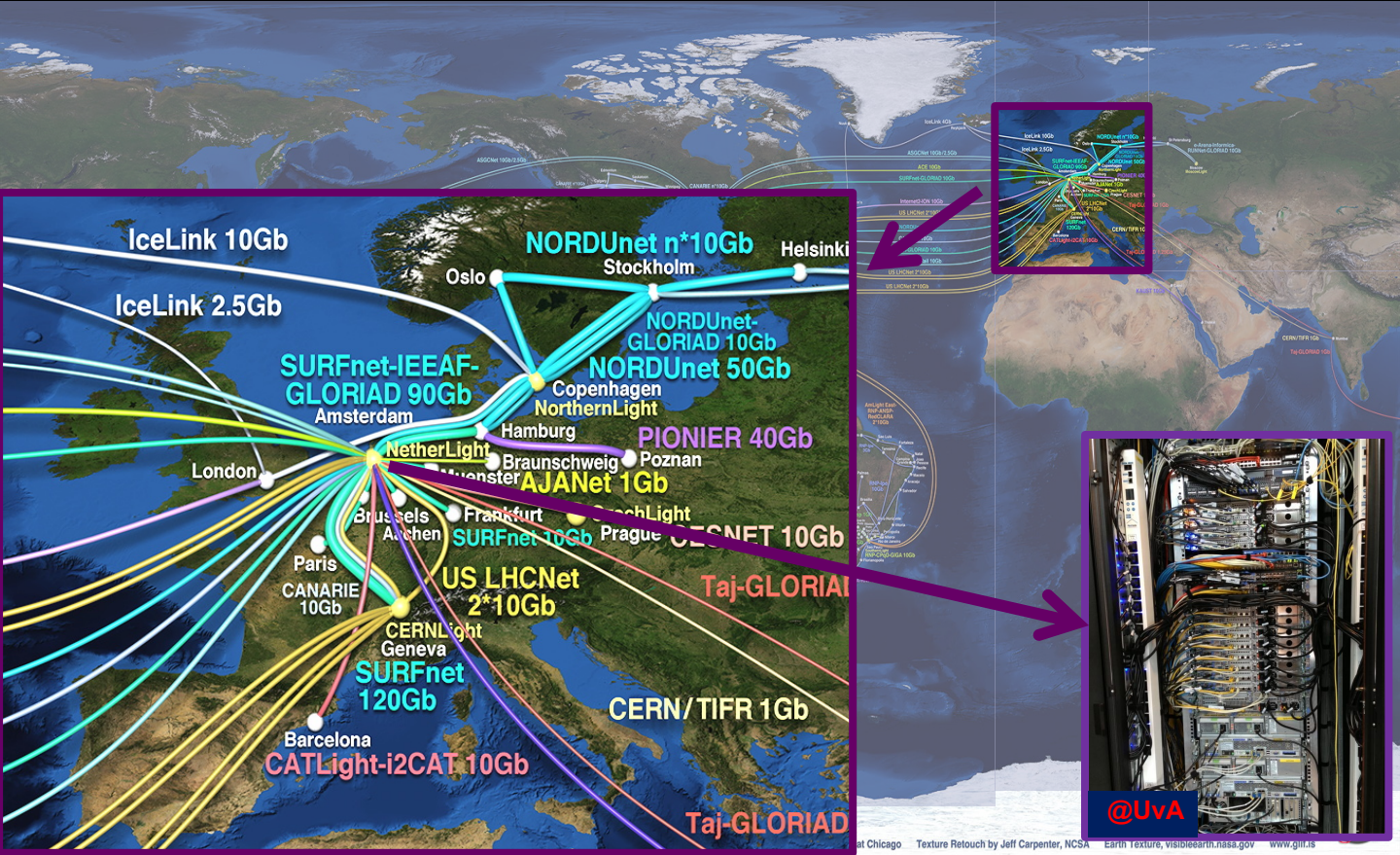
The GLIF – LightPaths around the World

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.



Amsterdam is a major hub in The GLIF

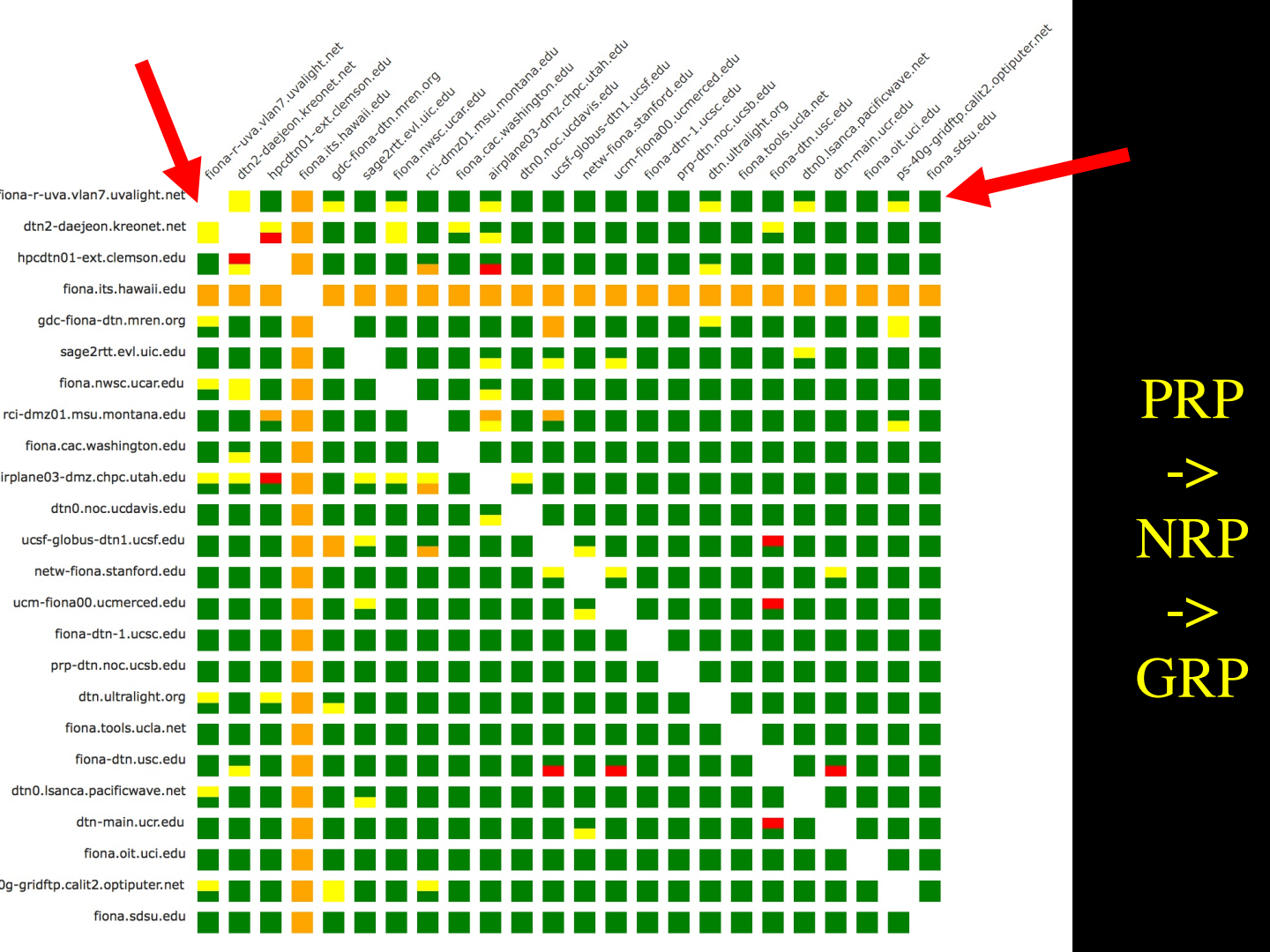
F Dijkstra, J van der Ham, P Grosso, C de Laat, "A path finding implementation for multi-layer networks", Future Generation Computer Systems 25 (2), 142-146.



Yesterday's Media Transport Method on the KL601 AMS-LAX-SAN!

8 TByte





PRP

->

NRP

->

GRP

Big Data: real time ICT for logistics

Data Logistics 4 Logistics Data (dl4ld)

Jan Burgmeijer, TNO, PI
Cees de Laat, UvA, Co-PI
Leon Gommans, KLM

Project Lead:

TNO



e:ofenedex

ORACLE

 **simacan**

Scientific Lead:



Partners:

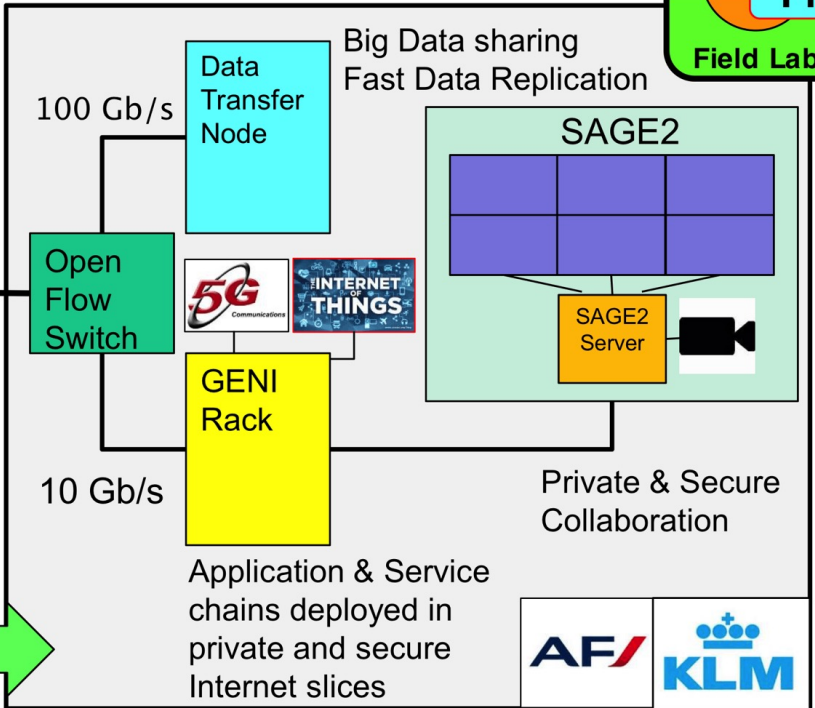
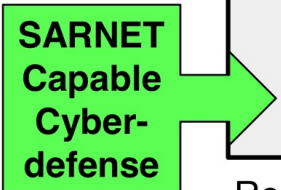
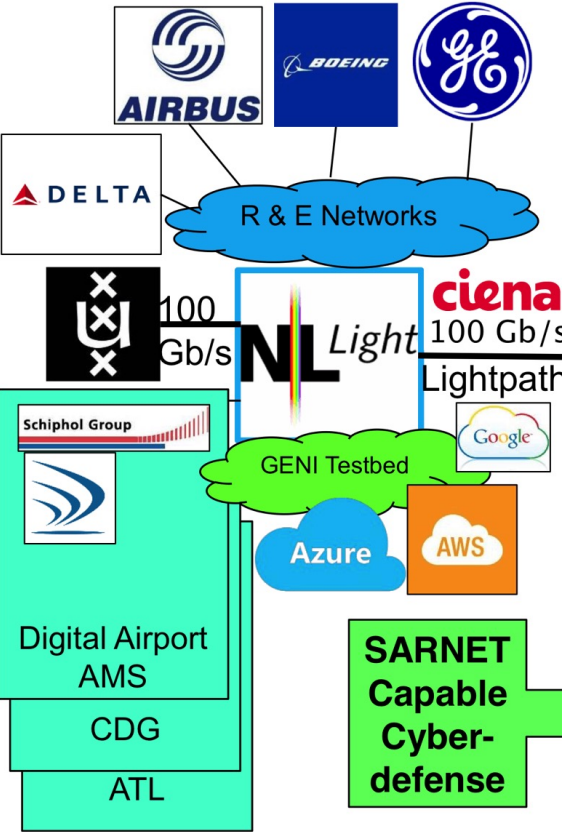
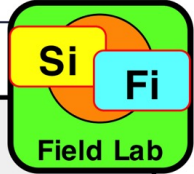
ciena

AIR FRANCE KLM

**X Gemeente
X Amsterdam
X**

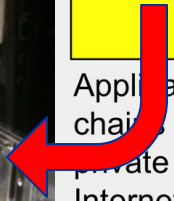
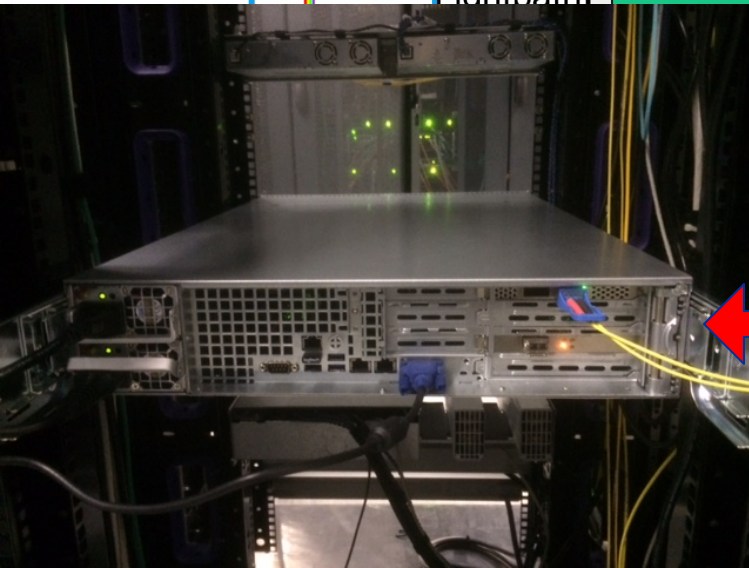
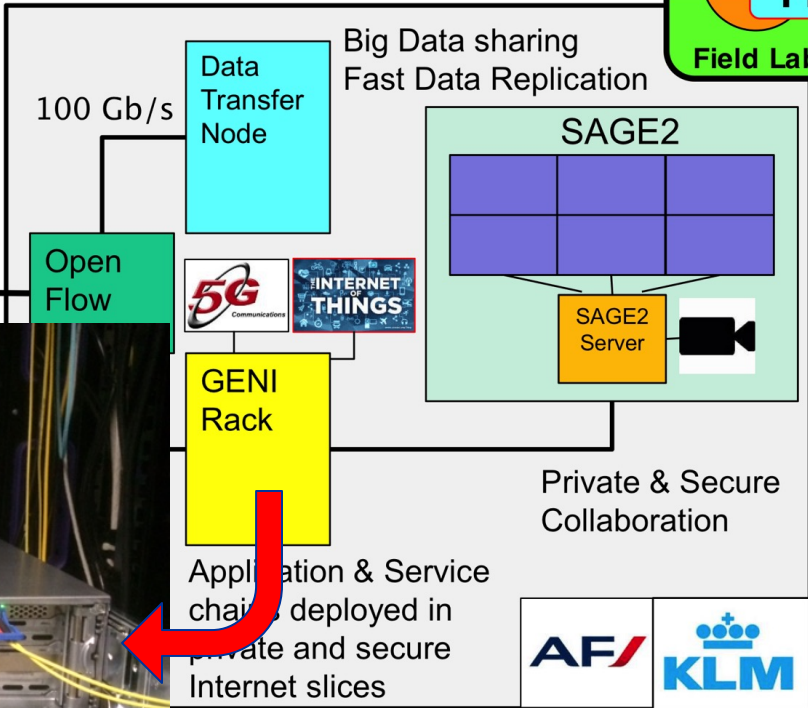
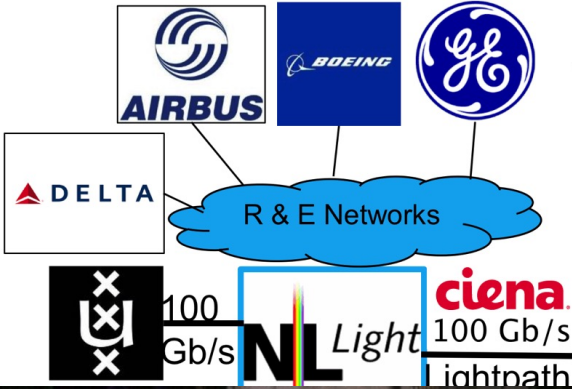
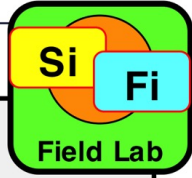
THALES
TRANSFIDES

Ambition to put capabilities into fieldlab



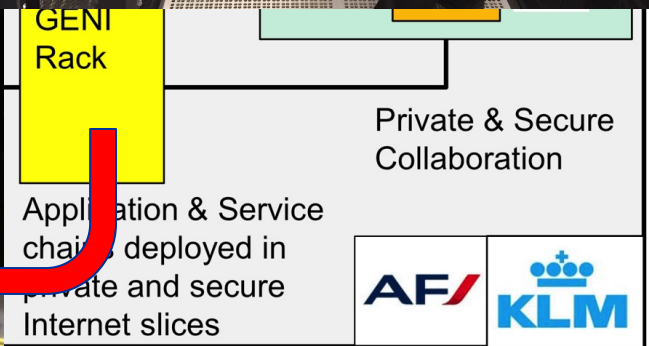
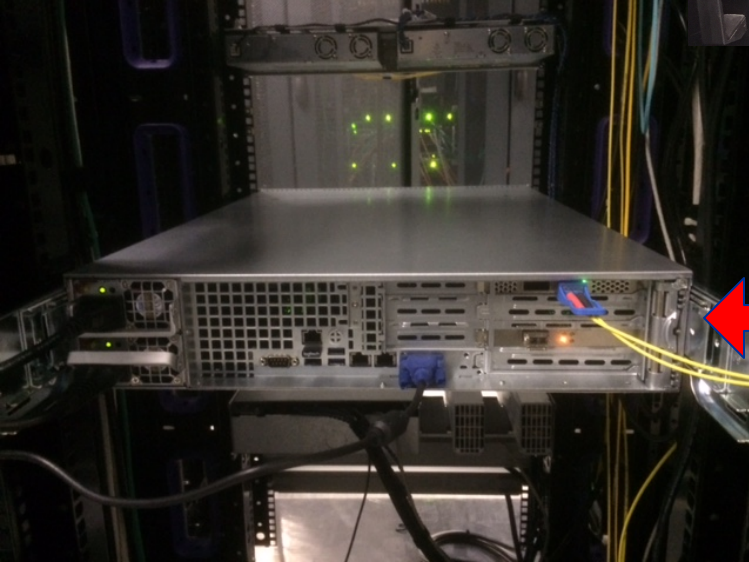
Re-enforcing ICT preconditions:
Each envisaged site has similar elements

Ambition to put capabilities into fieldlab



ing ICT preconditions:
saged site has similar elements





ing ICT preconditions:
saged site has similar elements

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

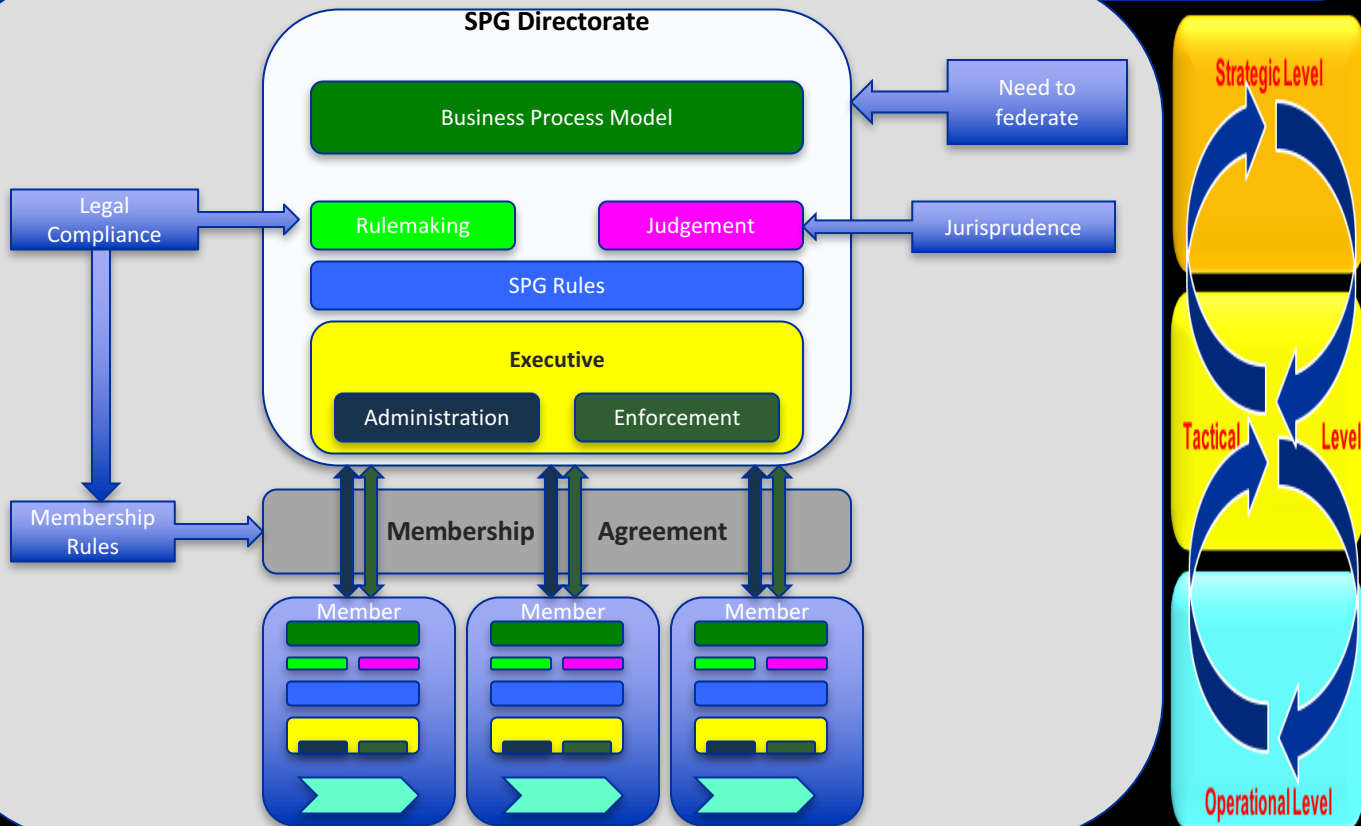


Main problem statement

- 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?



Observe an Alliance as a Service Provider Group system in terms of risk, cost & benefits



Approach

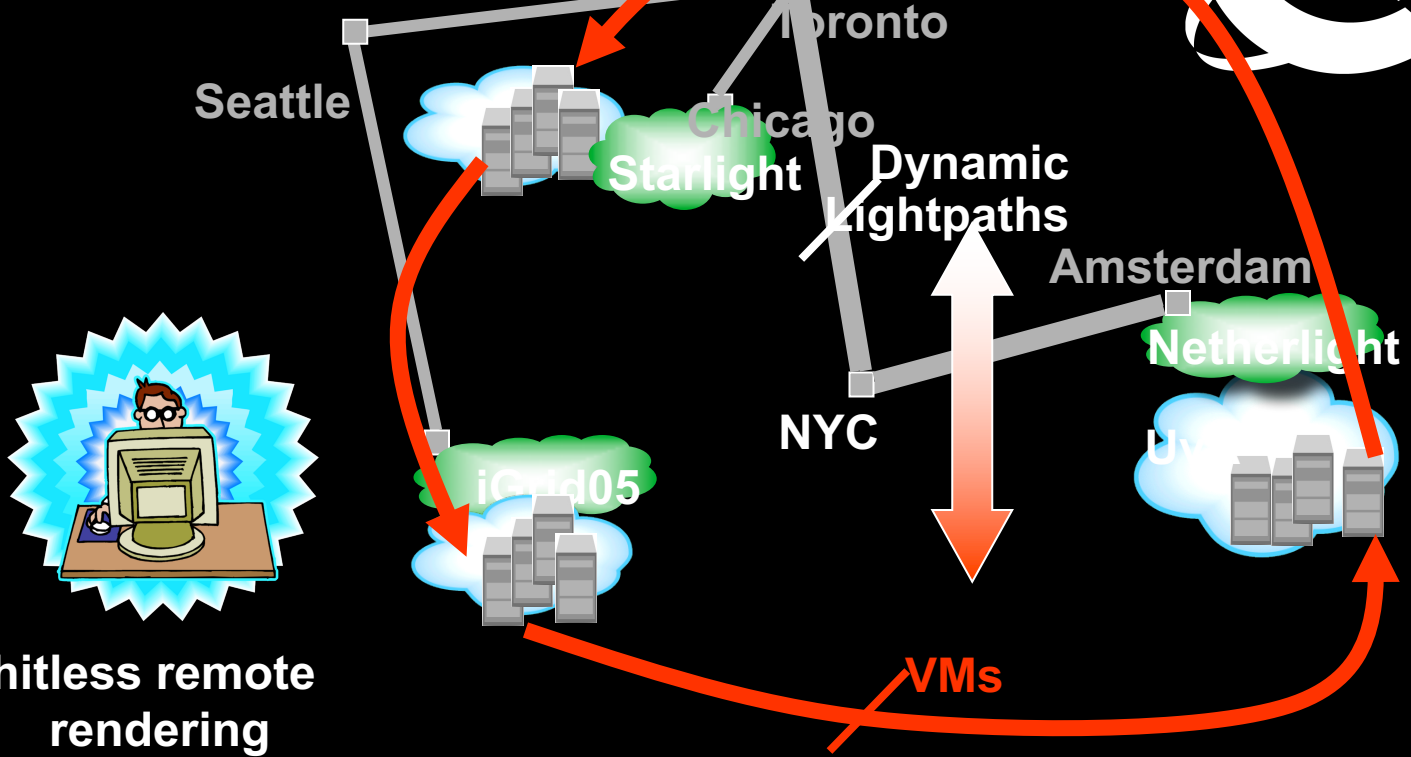
- Strategic:
 - Translate legislation into machine readable policy
 - Define data use policy
 - Trust evaluation models & metrics
- Tactical:
 - Map app given rules & policy & data and resources
 - Bring computing and data to (un)trusted third party
 - Resilience
- Operational:
 - TPM & Encryption schemes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



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 igrd2005!



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

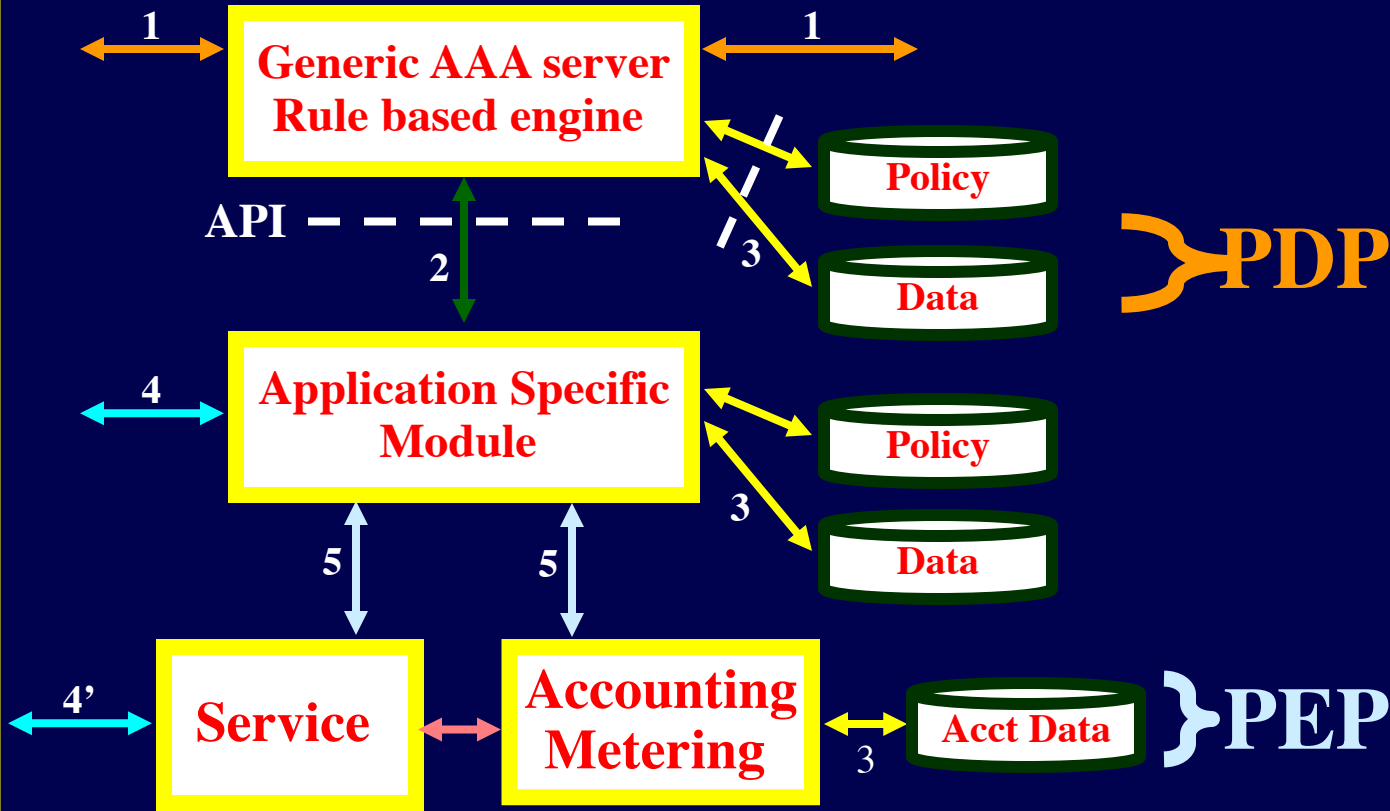
- San Diego – Amsterdam, 1GE, RTT = 200 msec, downtime ≤ 1 sec

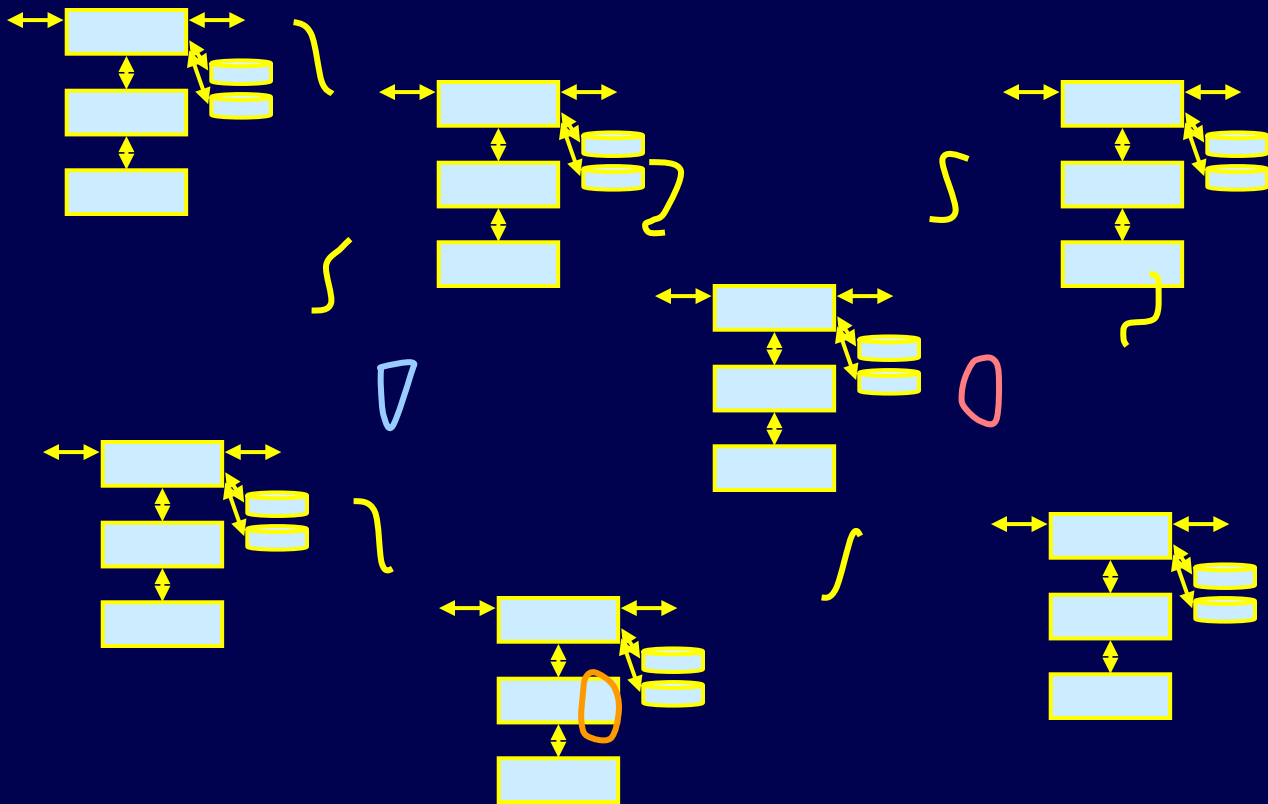
- Back to back, 1GE, RTT = 0.2-0.5 msec, downtime = ~ 0.2 sec*

**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





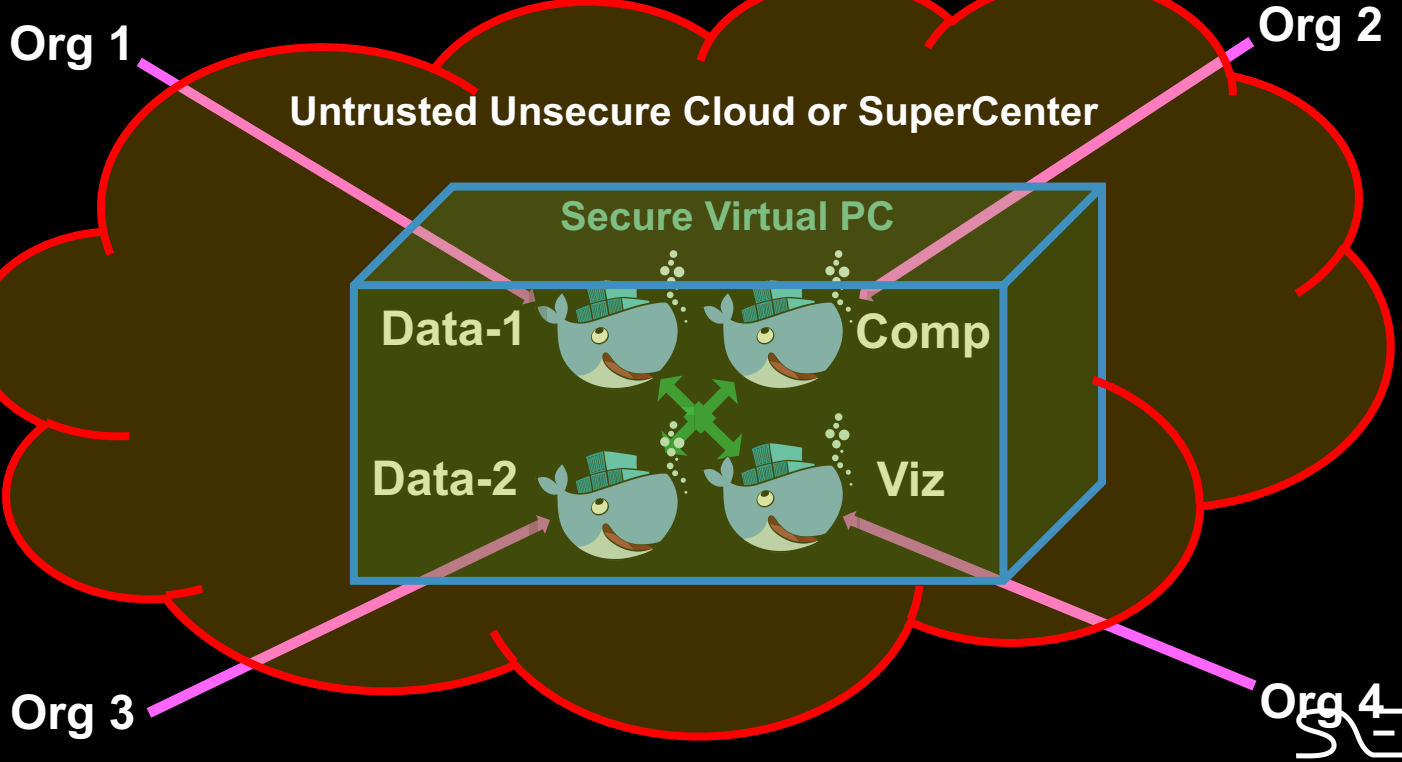
Line of research

- 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

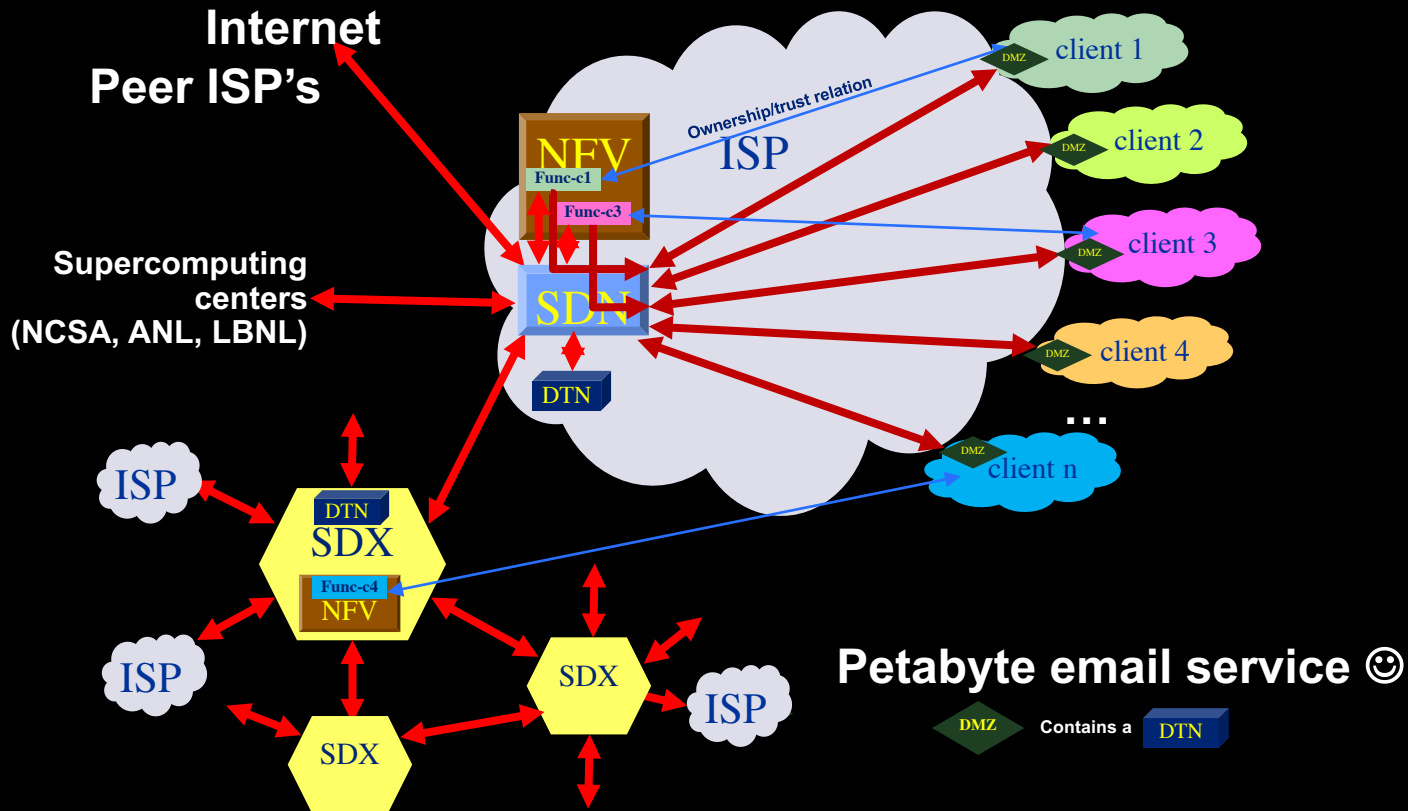
Secure Policy Enforced Data Processing



- 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



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



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

The proposed research focuses on:

- the design and implementation of capabilities that are required for creating value-rich logistic information services
- creation of innovative solutions that allow stakeholders to agree on how data is stored, accessed, shared and transformed in a controllable, enforceable, accountable, auditable and goal-oriented fashion.

Q & A

TNO



evofenedex

ORACLE

 **simacan**

More info: <http://delaat.net/dl4ld>
<http://delaat.net/sarnet>



ciena

AIR FRANCE KLM

**X Gemeente
X Amsterdam
X**

**THALES
TRANSIDES**