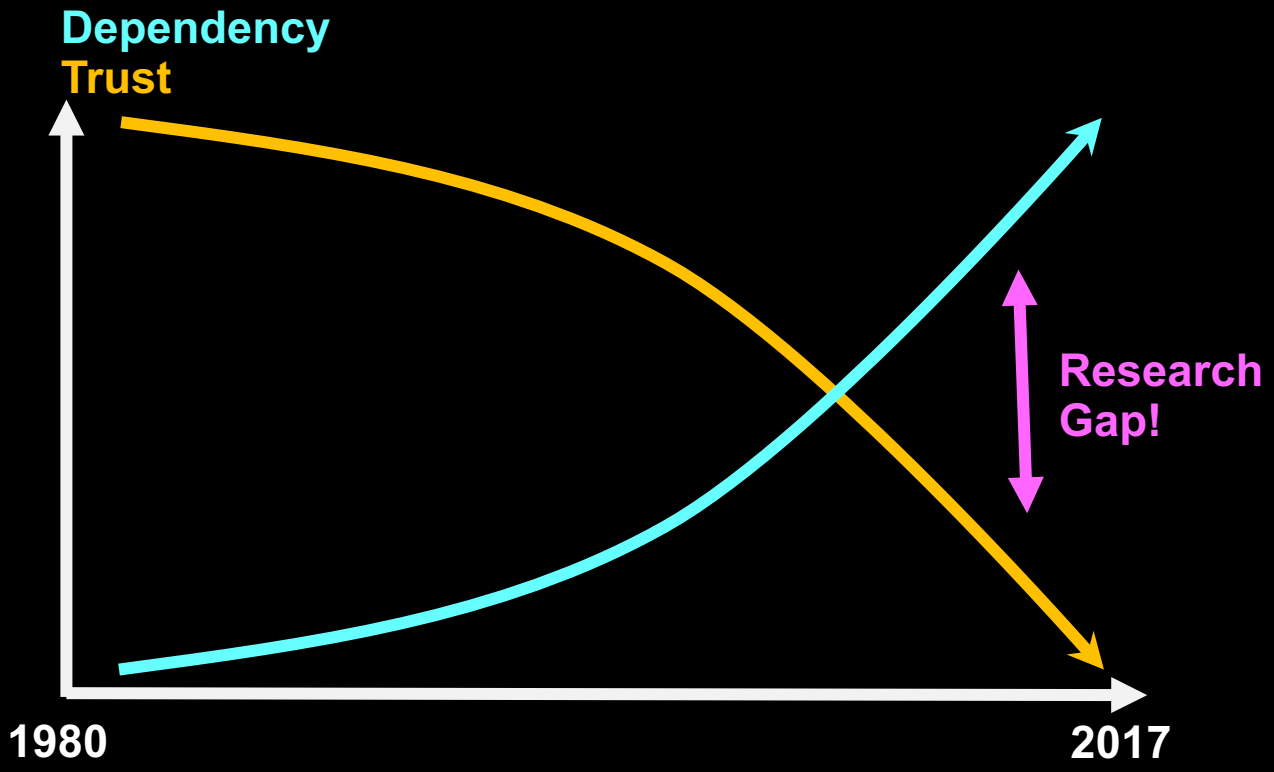
Distributed Big Data Assets Sharing & Processing

Big Data Hub infrastructure

Jan Wester, C. de Laat, L. Gommans

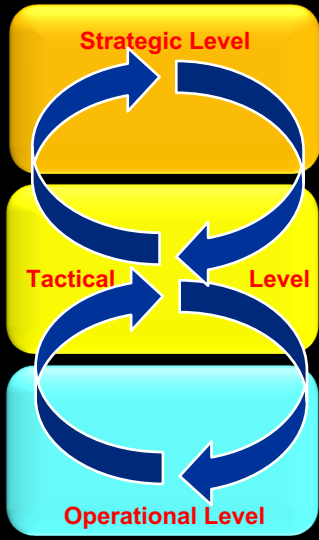
TNO
System & Network Engineering, University of Amsterdam
AirFrance KLM

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

National Scale

City / regional Scale

Campus / Enterprise Scale

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

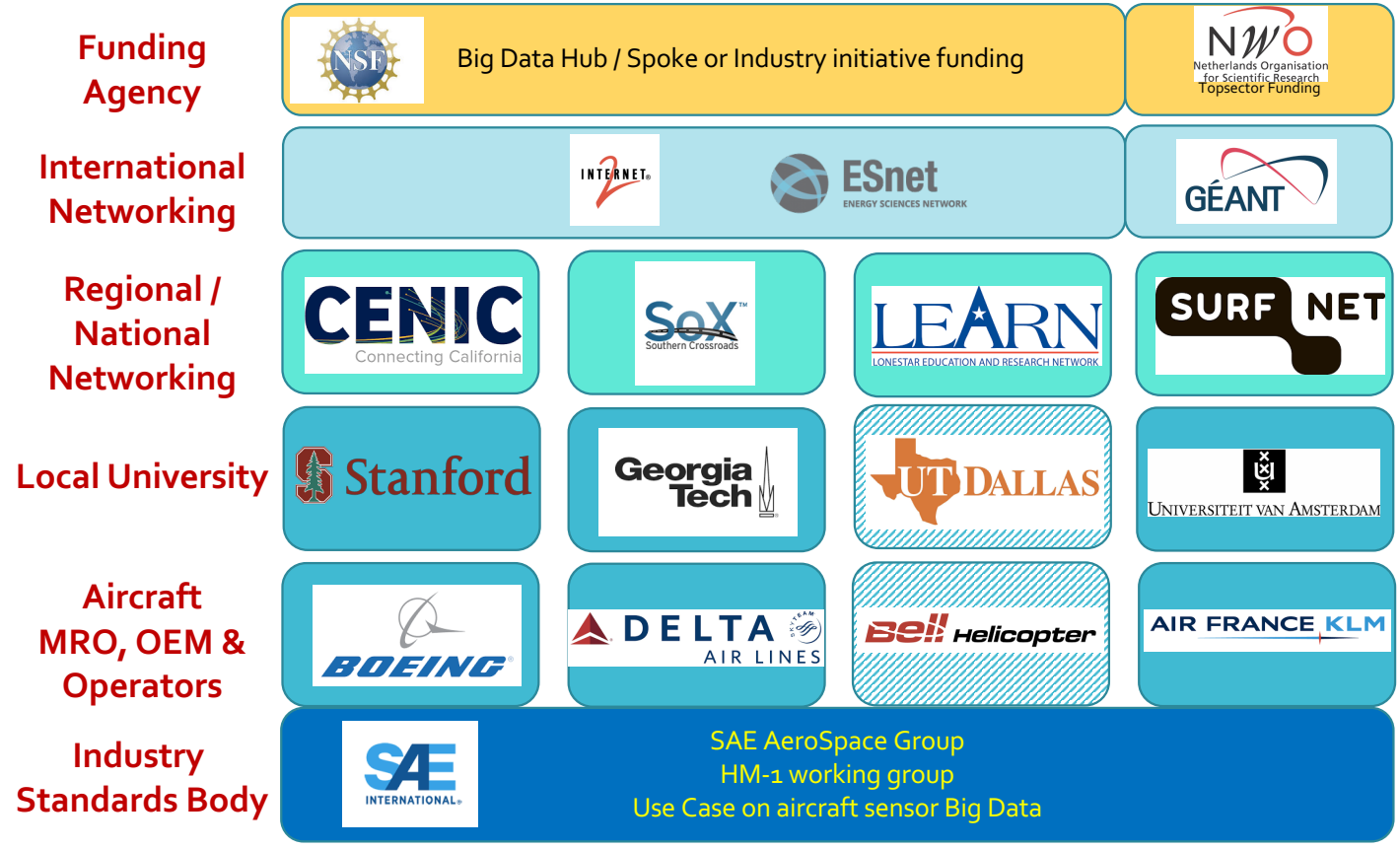
Cargo Logistics Data
(C1) DaL4LoD
(C2) Secure scalable policy-enforced distributed data Processing
(using blockchain)

NLIP iShare project

iSHARE
powered by NLIP

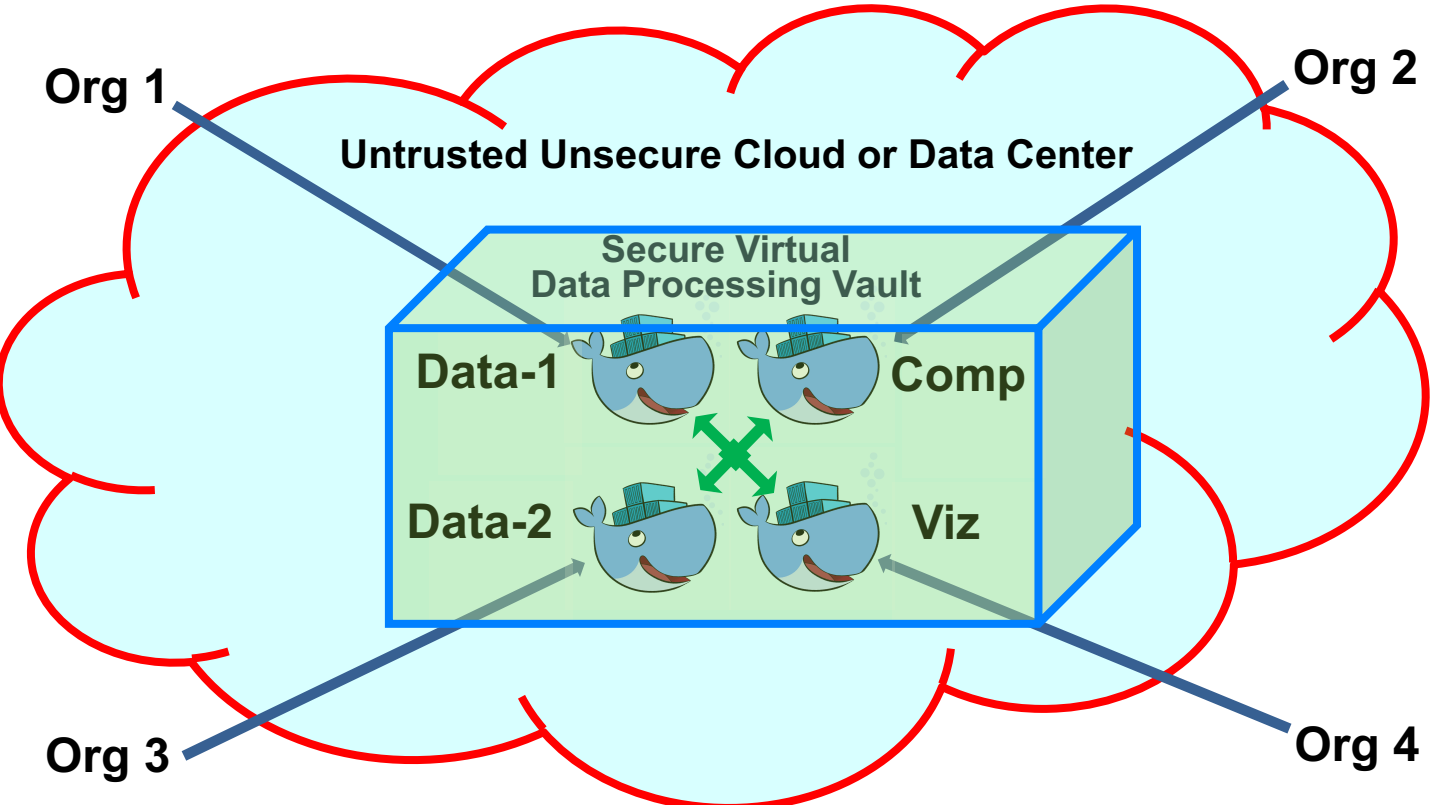
Cybersecurity Big Data
NWO COMMIT/
SARNET project
3.5 FTE

SAE Use Case envisaged research collaboration

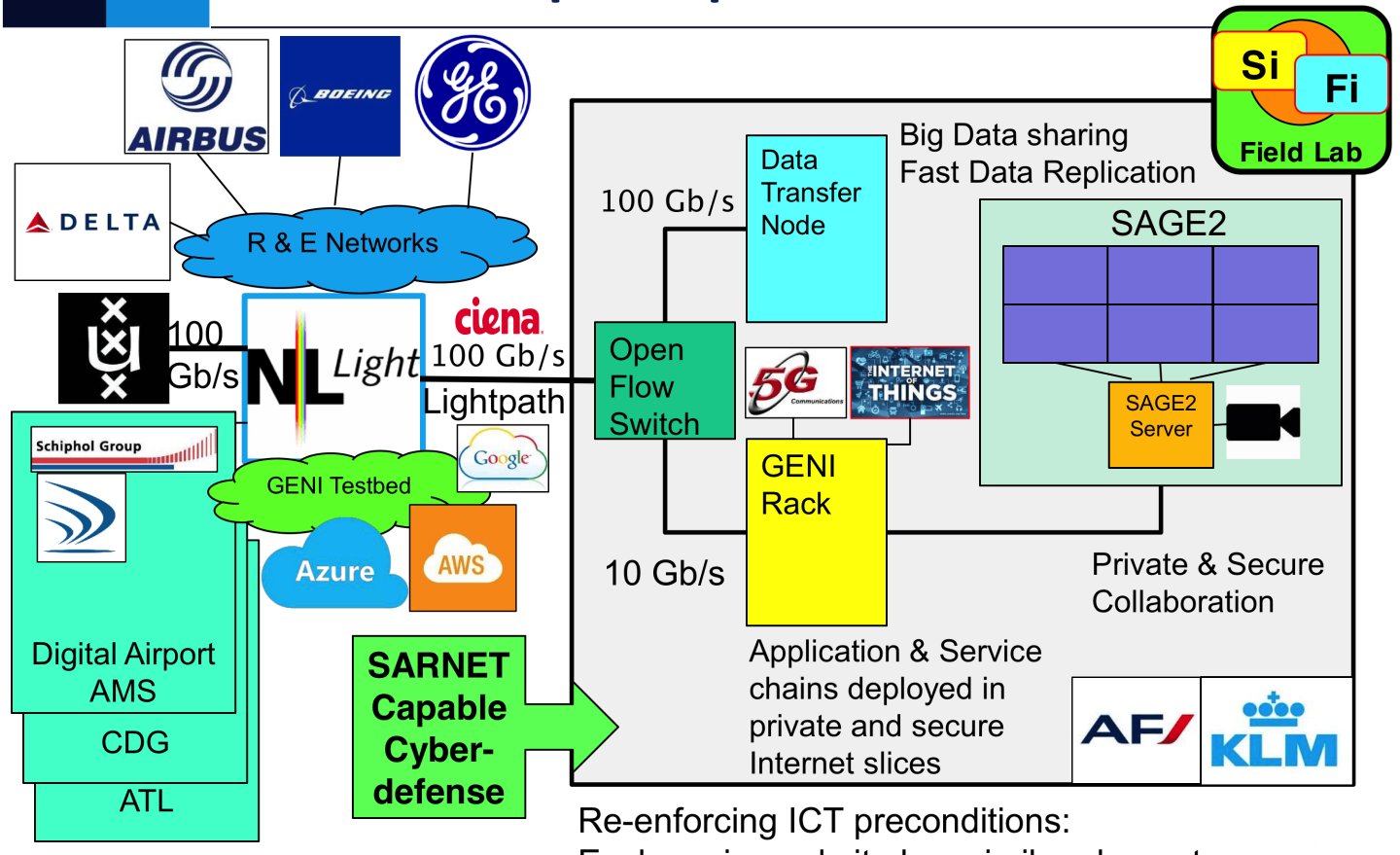


Example model: Policy Enforced Data Processing

- Bringing data and processing software from competing organizations 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)



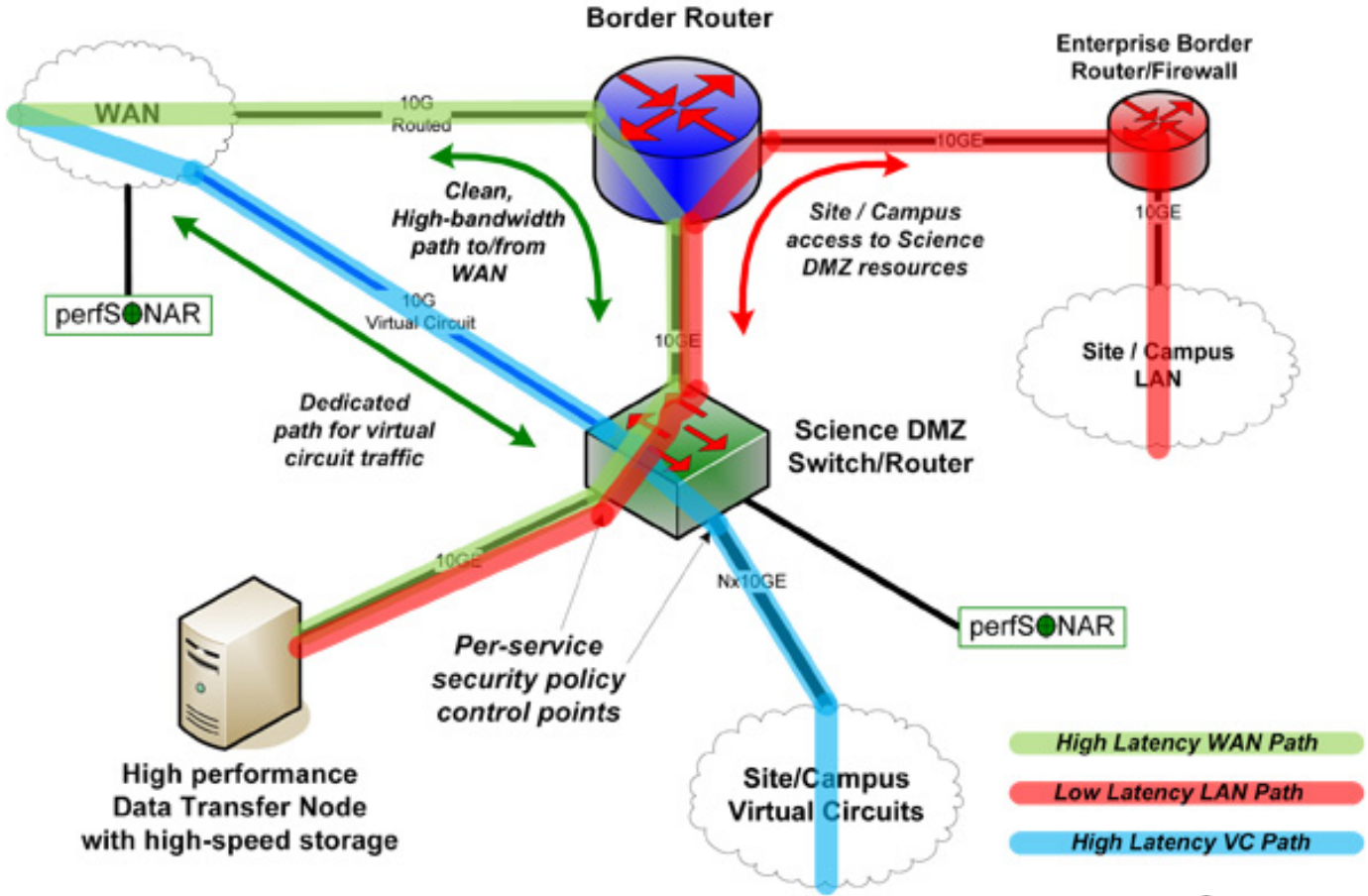
Ambition to put capabilities into fieldlab



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

AIRFRANCE KLM

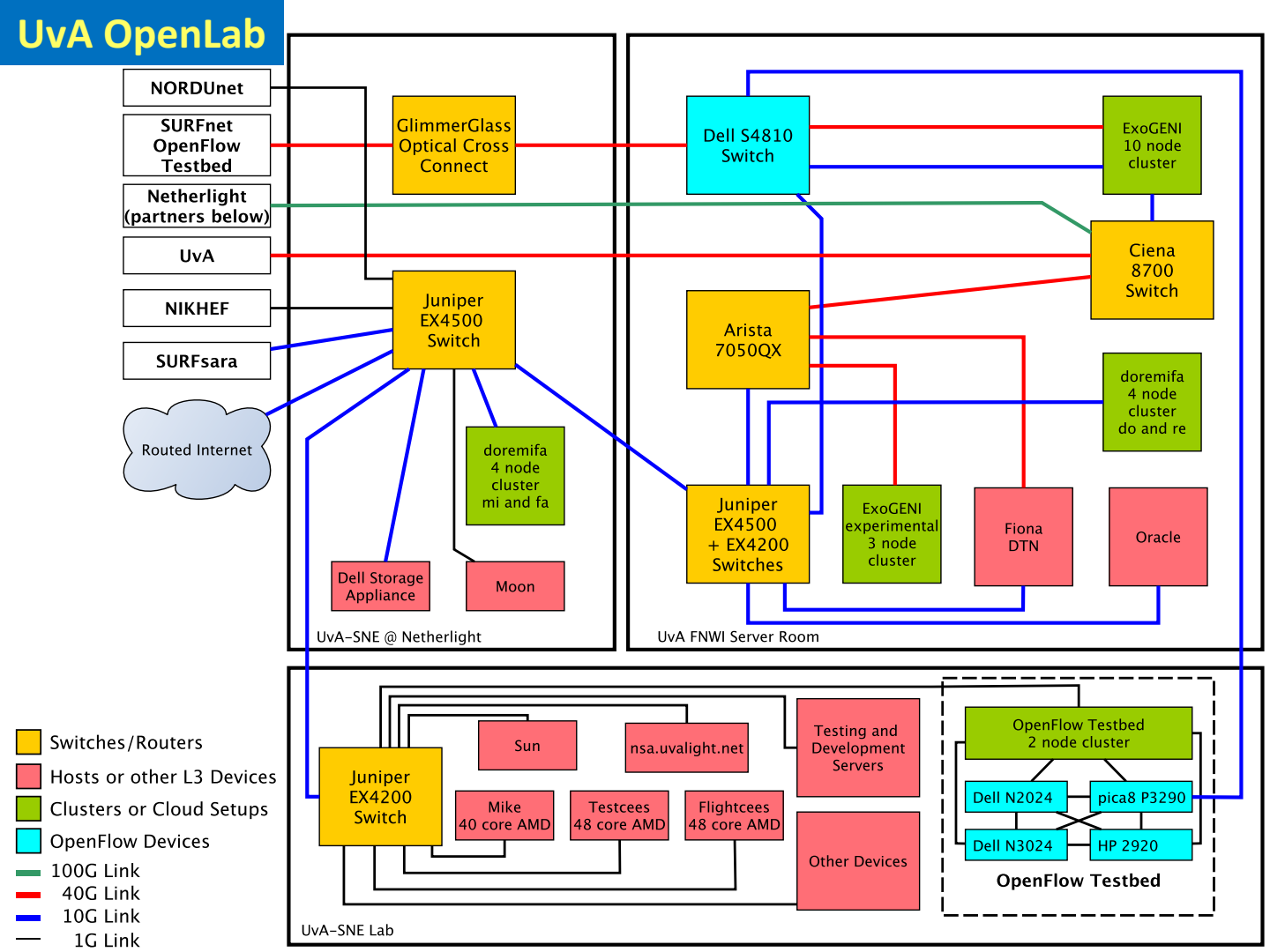
Science-DMZ



CENIC

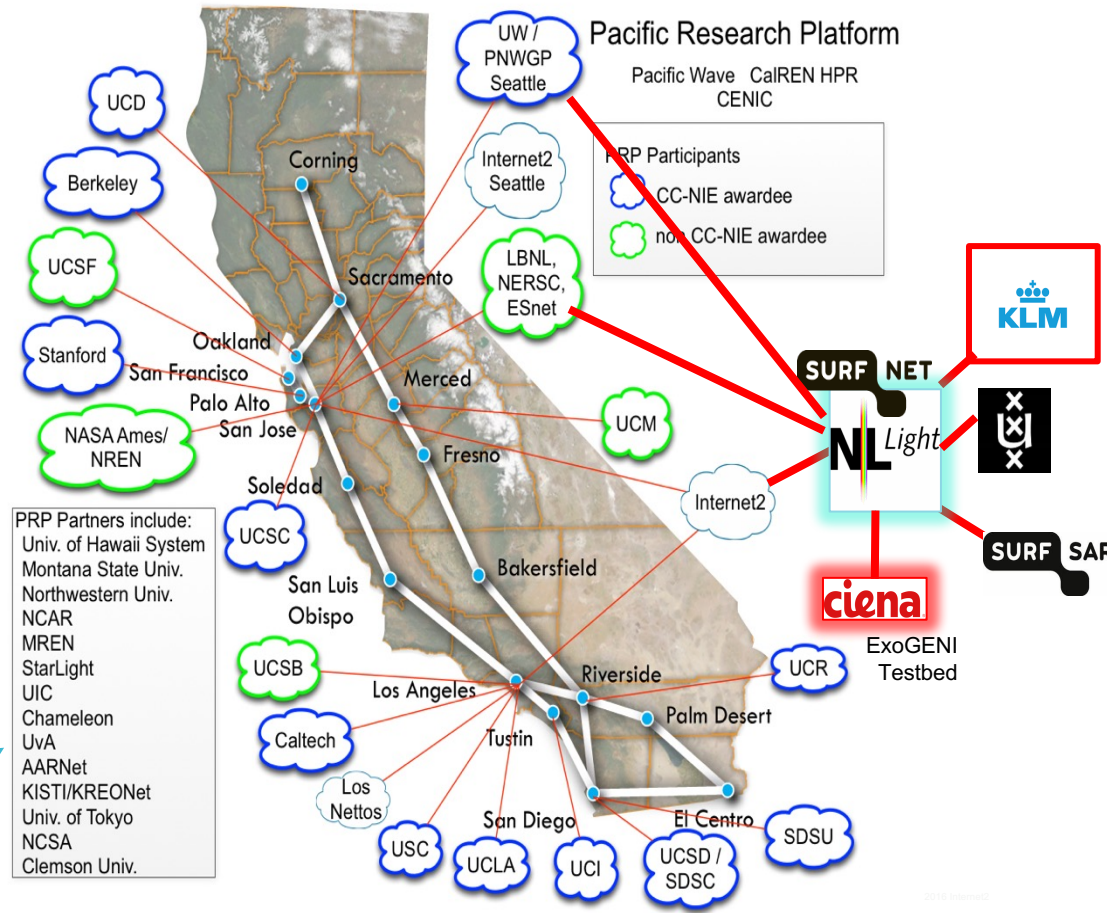
ESnet

UvA OpenLab


NORDUnet
SURFnet OpenFlow Testbed
Netherlight (partners below)
UvA
NIKHEF
SURFsara
Routed Internet
GlimmerGlass Optical Cross Connect
Juniper EX4500 Switch
doremifa 4 node cluster mi and fa
Dell Storage Appliance
Moon
UvA-SNE @ Netherlight
Dell S4810 Switch
ExoGENI 10 node cluster
Ciena 8700 Switch
Arista 7050QX
doremifa 4 node cluster do and re
Juniper EX4500 + EX4200 Switches
ExoGENI experimental 3 node cluster
Fiona DTN
Oracle
UvA FNWI Server Room
Juniper EX4200 Switch
Sun
nsa.uvalight.net
Mike 40 core AMD
Testcees 48 core AMD
Flightcees 48 core AMD
Testing and Development Servers
Other Devices
OpenFlow Testbed 2 node cluster
Dell N2024
pica8 P3290
Dell N3024
HP 2920
OpenFlow Testbed
UvA-SNE Lab
Switches/Routers
Hosts or other L3 Devices
Clusters or Cloud Setups
OpenFlow Devices
100G Link
40G Link
10G Link
1G Link

Pacific Research Platform testbed involvement

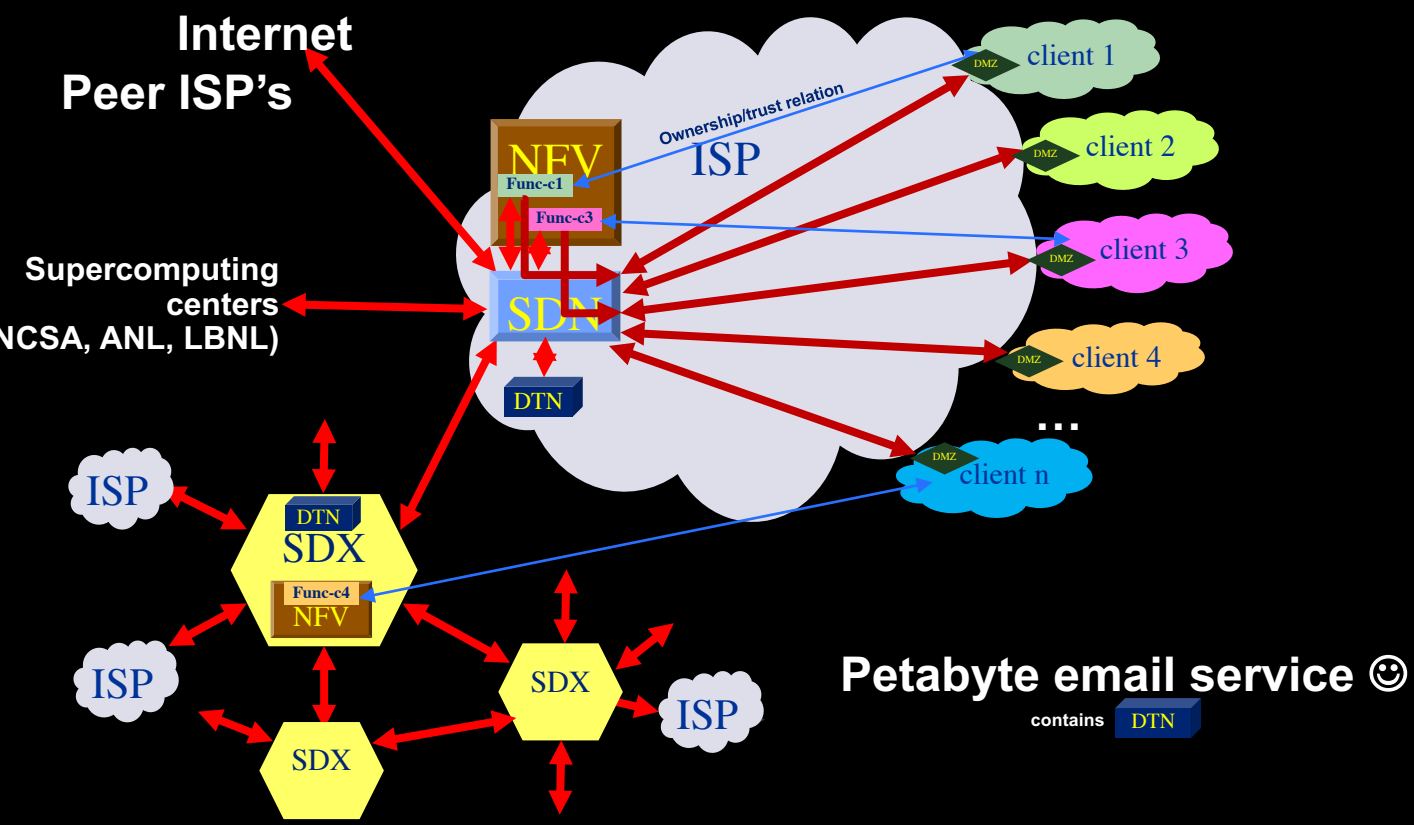
Research goal: Explore value of academic network research capabilities that enable innovative ways & models to share big data assets



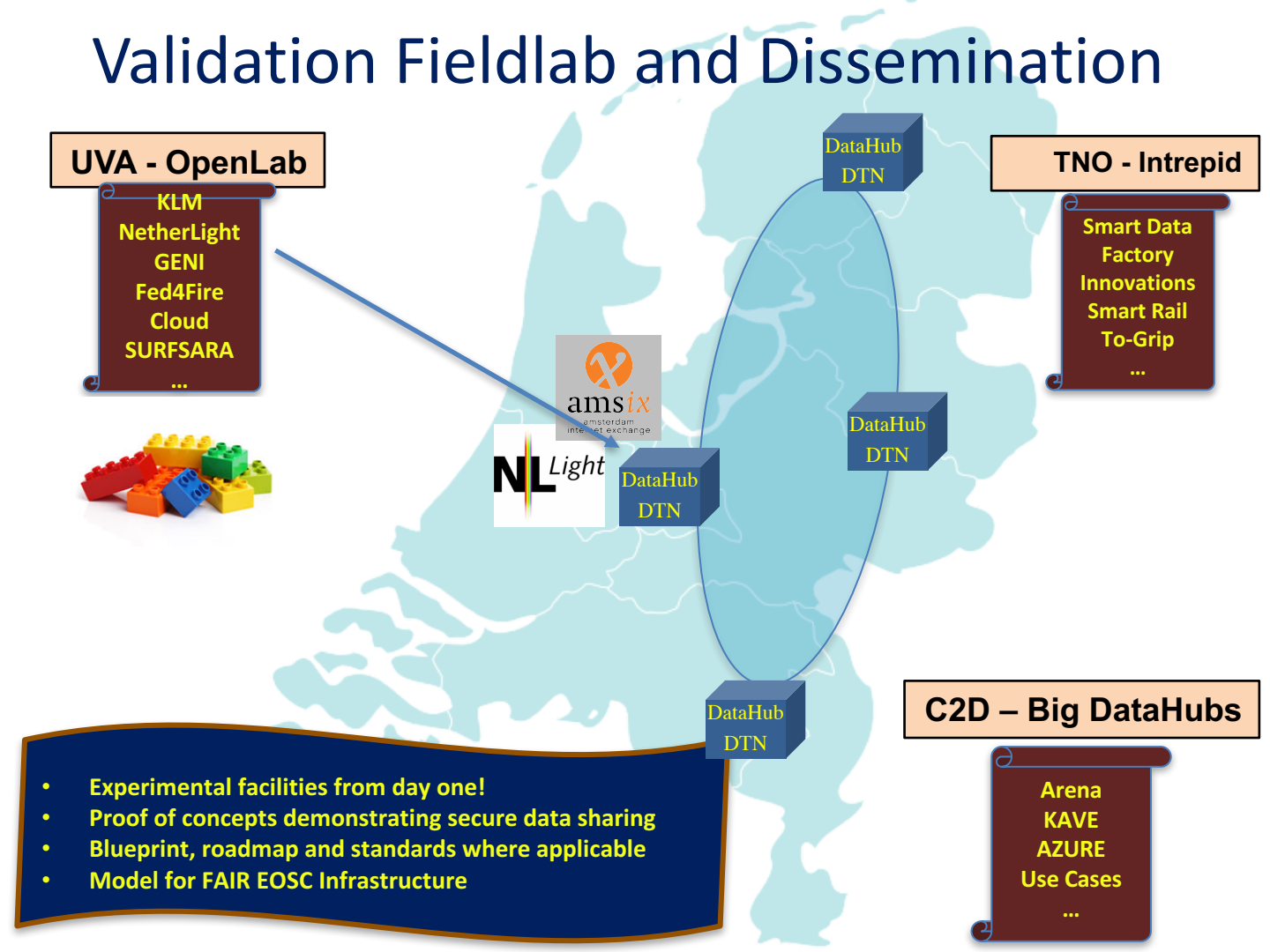
PRP PACIFIC RESEARCH PLATFORM

Note: this diagram represents a subset of sites and connections.

Networks of ScienceDMZ's & SDX's



Validation Fieldlab and Dissemination



- Experimental facilities from day one!
- Proof of concepts demonstrating secure data sharing
- Blueprint, roadmap and standards where applicable
- Model for FAIR EOSC Infrastructure

Q&A

Program at Global Summit I2 in Washington DC April 2017:

15h00 Cees de Laat, University of Amsterdam
Trusted Data Processing in Untrusted Environments.

15h05 Leon Gommans, Air France KLM
Trusted Big Data Sharing.

15h25 Rodney Wilson
Programmable Supernetworks, Science DMZ based Networking.

15h30 Panel of stakeholders Flash talks (~3 min each):
Inder Monga - ESnet - Data Science Driving Discovery.
Matt Zekauskas - Internet2 - Thoughts on Internet2 and Trusted Large Data Transfer.
Jerry Sobieski - NORDUnet - Issues of Big Data Sharing in a Global Science Collaboration.
Adam Slagell – NCSA - What are we trusting?

15h45 Panel discussion moderated by Cees de Laat

16h00 End of session.