Digital Data Markets: real time ICT for logistics
Data Logistics 4 Logistics Data (dl4ld)

- PI's: prof.dr. Robert Meijer (TNO & UvA), prof.dr.ir. Cees de Laat (UVA)
- PL: dr.ir. Harrie Bastiaansen
- TNO: dr. Wout Hofman, dr. Ir. Anne Fleur van Veenstra, Simon Dalmolen MSc
- UvA: dr. Paola Grosso, prof.dr. Tom van Engers
- KLM & UvA: dr. ing. Leon Gommans
- KPMG & UvA: prof. dr. Sander Klous
- Thales Nederland: dr. Kees Nieuwenhuis
- CIENA: Rodney Wilson, Marc Lyonais
- ORACLE: Loek Hassing

TNO

ZZ

exofenedex

ORACLE

DL4LD

NWO
Netherlands Organisation for Scientific Research

ciena

AIR FRANCE KLM

Gemeente Amsterdam

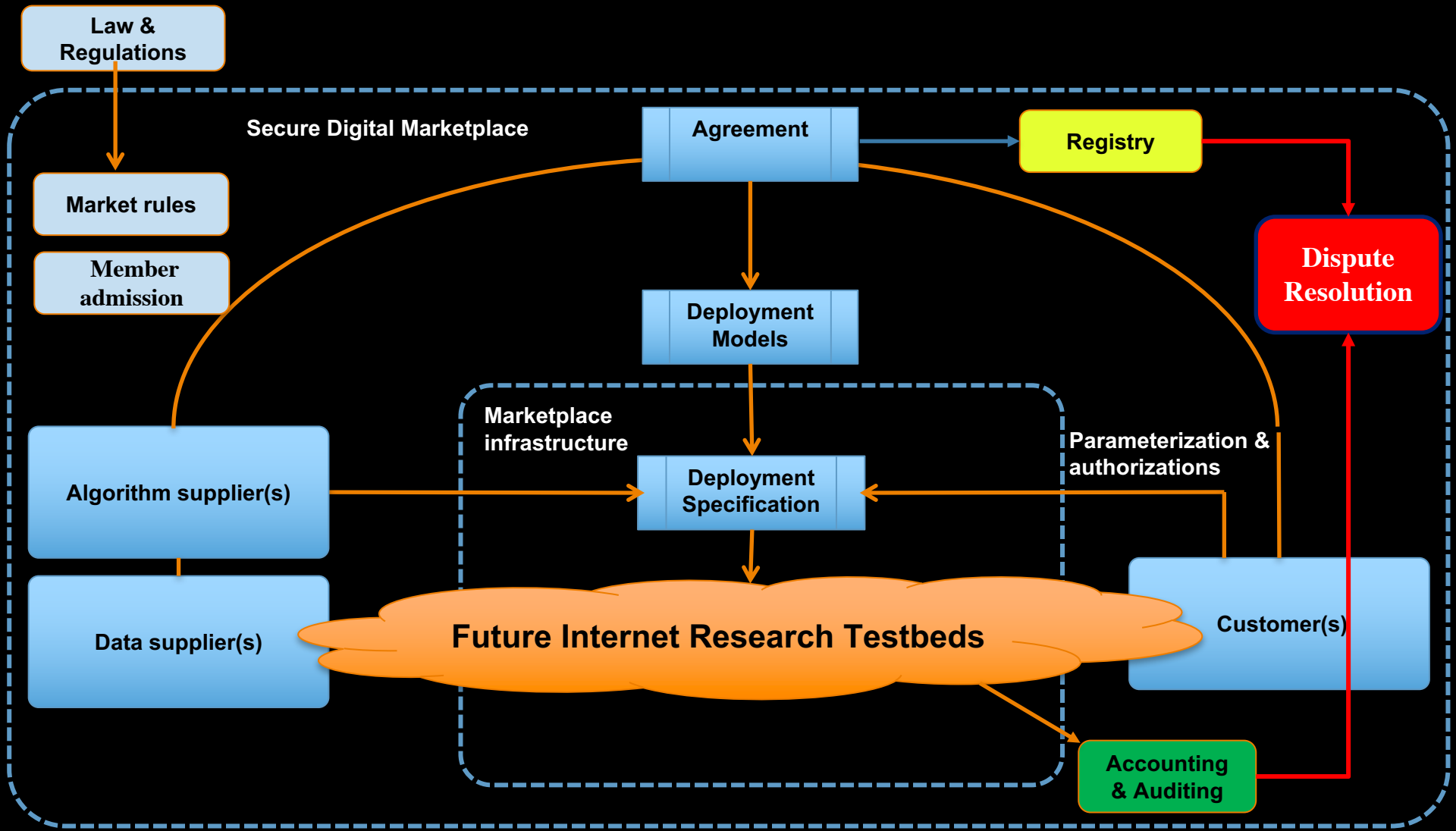
THALES

TRANSFIDES

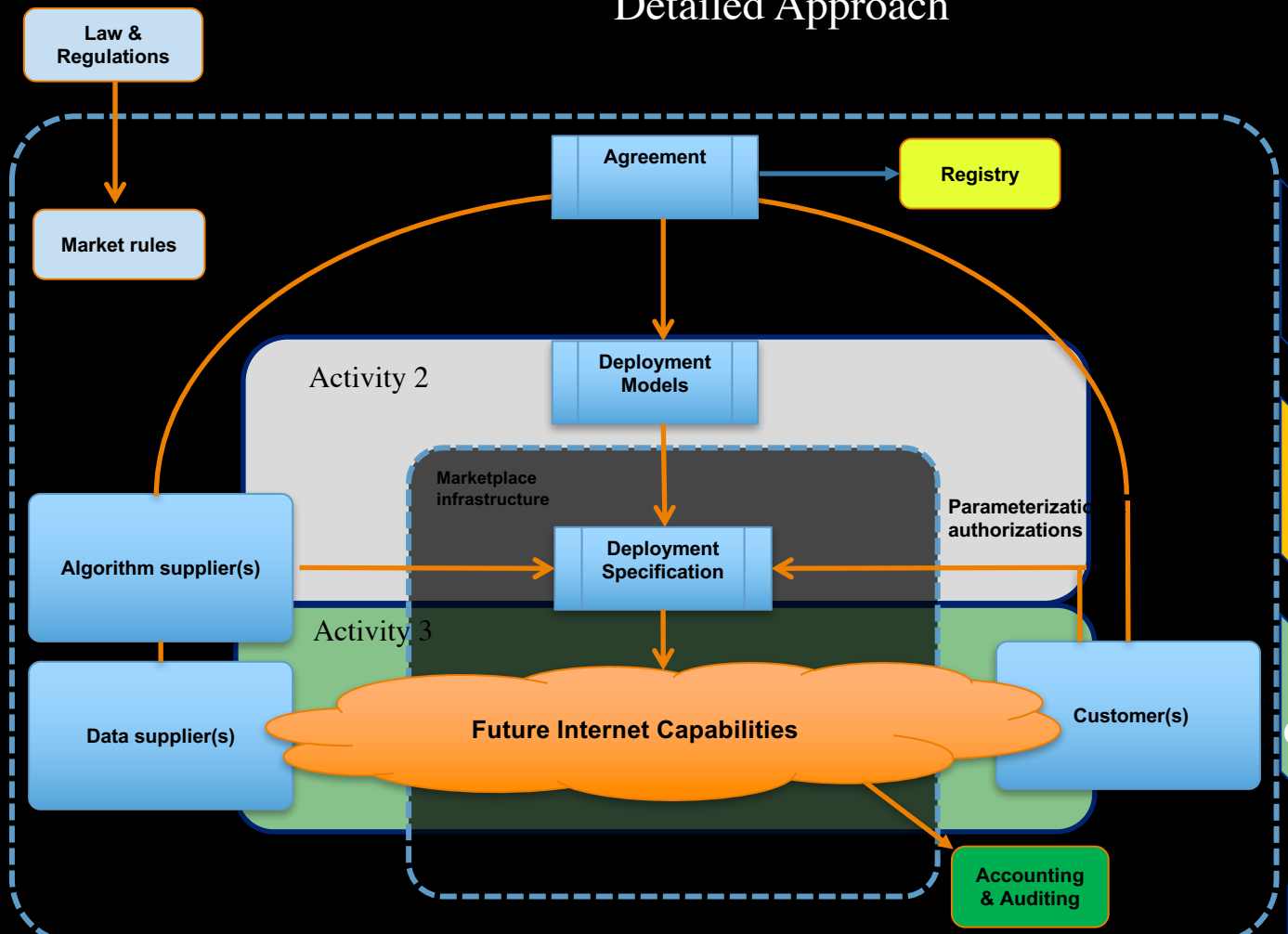
Main problem statement

- 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



Why & What

Strategic

- Digital business agreement negotiation & trust organized by member organization
- Architecture driven deployment of trusted system functions needed by business
- Creating Inter-organizational solutions

How

Tactical

- Digital realization & implementation of trust
- Creating archetypes of deployment models.
- Semantic composition of services
- Incorporating forward looking infrastructure architectures

With what

Operational

- Inter-organization service quality & consistency in practice
- Multidomain aspects (identity, AAA)
- Infrastructure programmability
- Programmable component deployment

National & International Academic Fieldlabs

WP4 Knowledge Exchange

WP3 Sector Requirements

Topsector Fieldlabs (Logistics, Smart Industry, Health, AgriFood ...)

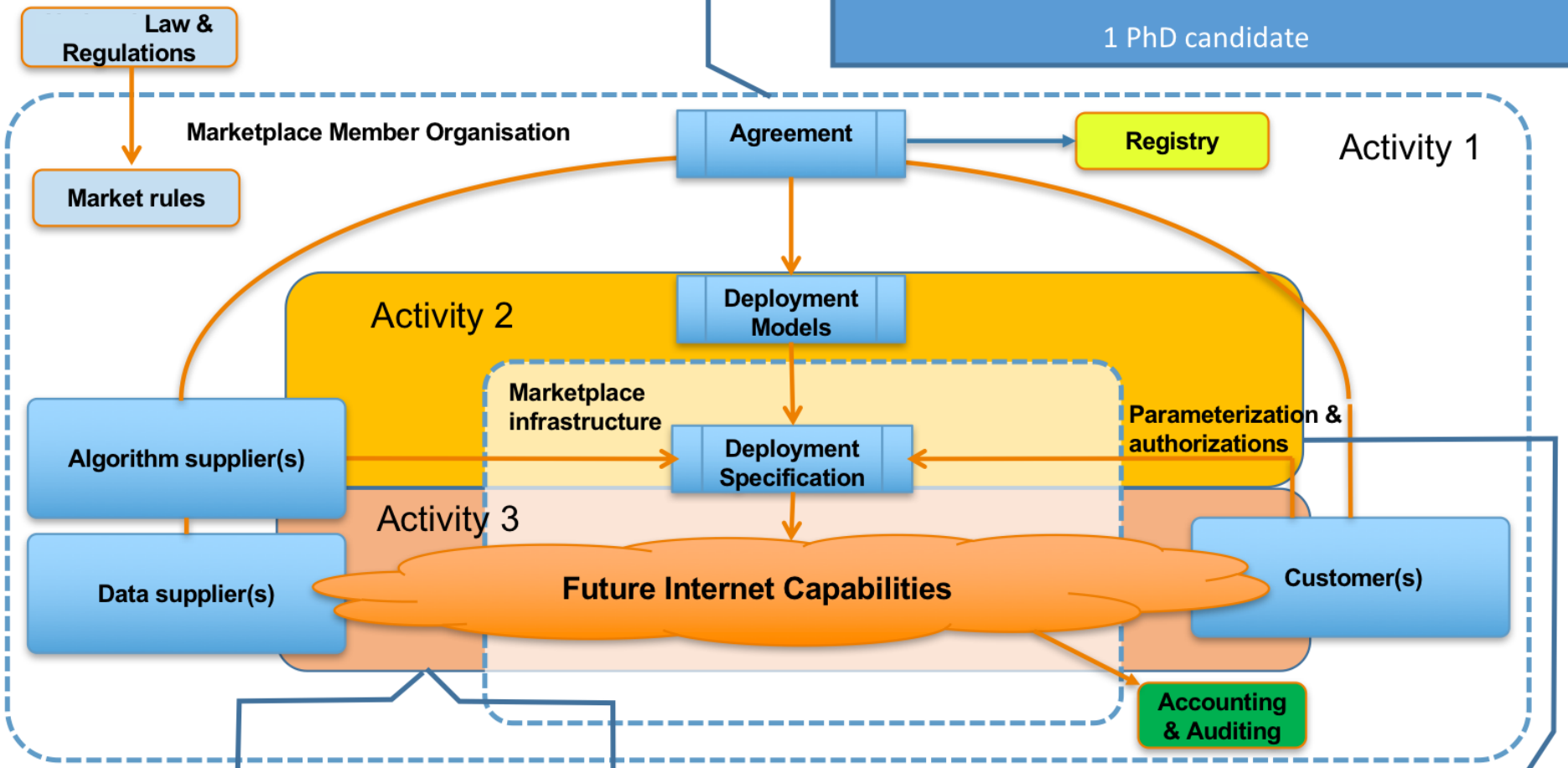
WP3 Blue print

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

1 Prgmr

Openlab, KLM, Ciena, GLIF
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

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

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 + PEP technology

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

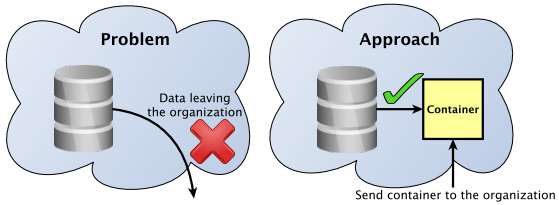
Universiteit van Amsterdam

Łukasz Makowski, Daniel Romão, Cees de Laat, Paola Grosso
System and Networking Research Group, University of Amsterdam

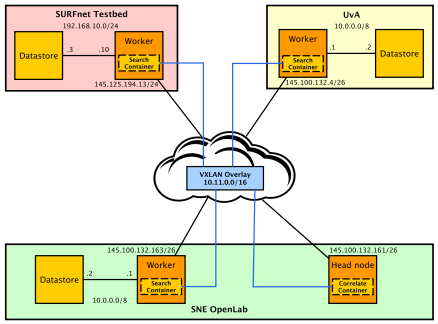
System and Network Engineering

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

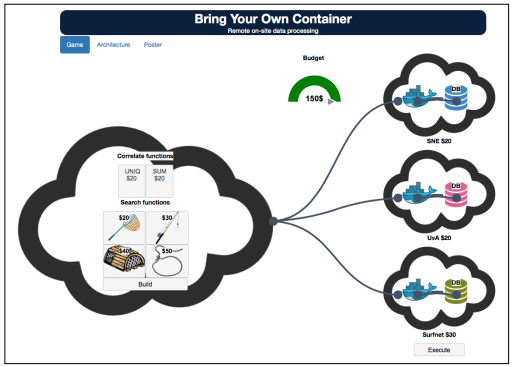
Our SC16 demo is a gamification of the remote dataset processing architecture.

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

SURF NET

The VM Turntable Demonstrator

hitless remote rendering

Seattle

Toronto

Chicago

Starlight

Dynamic Lightpaths

Amsterdam

Netherlight

NYC

Grid05

VMs

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 <= 1 sec
- Back to back, 1GE, RTT = 0.2-0.5 msec, downtime = ~0.2 sec*
- **Clark et al. NSDI 05 paper. Different workloads*

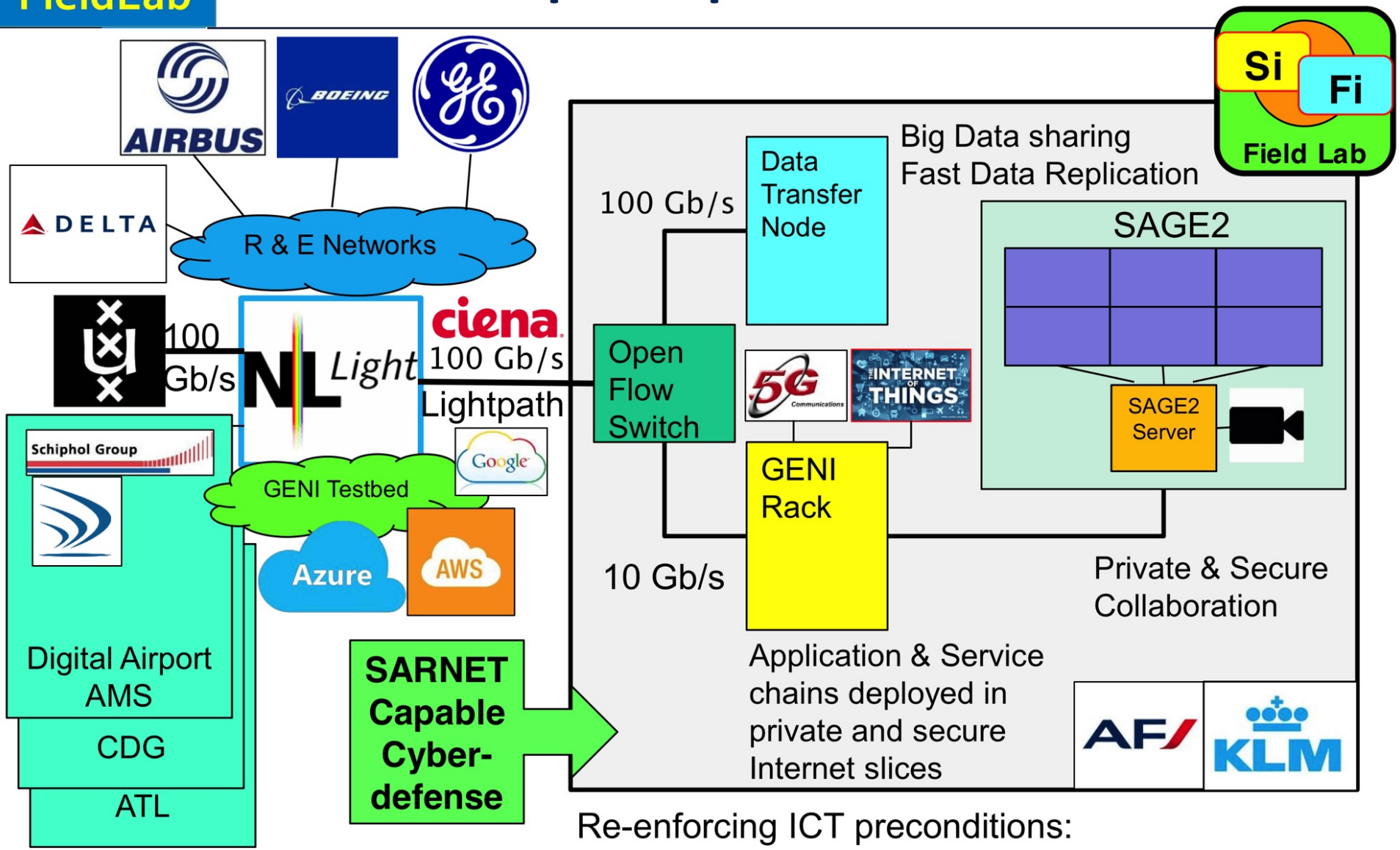
VM + Lightpaths across MAN/WAN are deemed a powerful and general alternative to RPC, GRAM approaches

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

F. Travostino, P. Daspit, L. Gommans, C. Jog, C.T.A.M. de Laat, J. Mambretti, I. Monga, B. van Oudenaarde, S. Raghunath and P.Y. Wang, "Seamless Live Migration of Virtual Machines over the MAN/WAN", Future Generation Computer Systems, Volume 22, Issue 8, October 2006, Pages 901-907.

AF/KLM FieldLab

Ambition to put capabilities into fieldlab

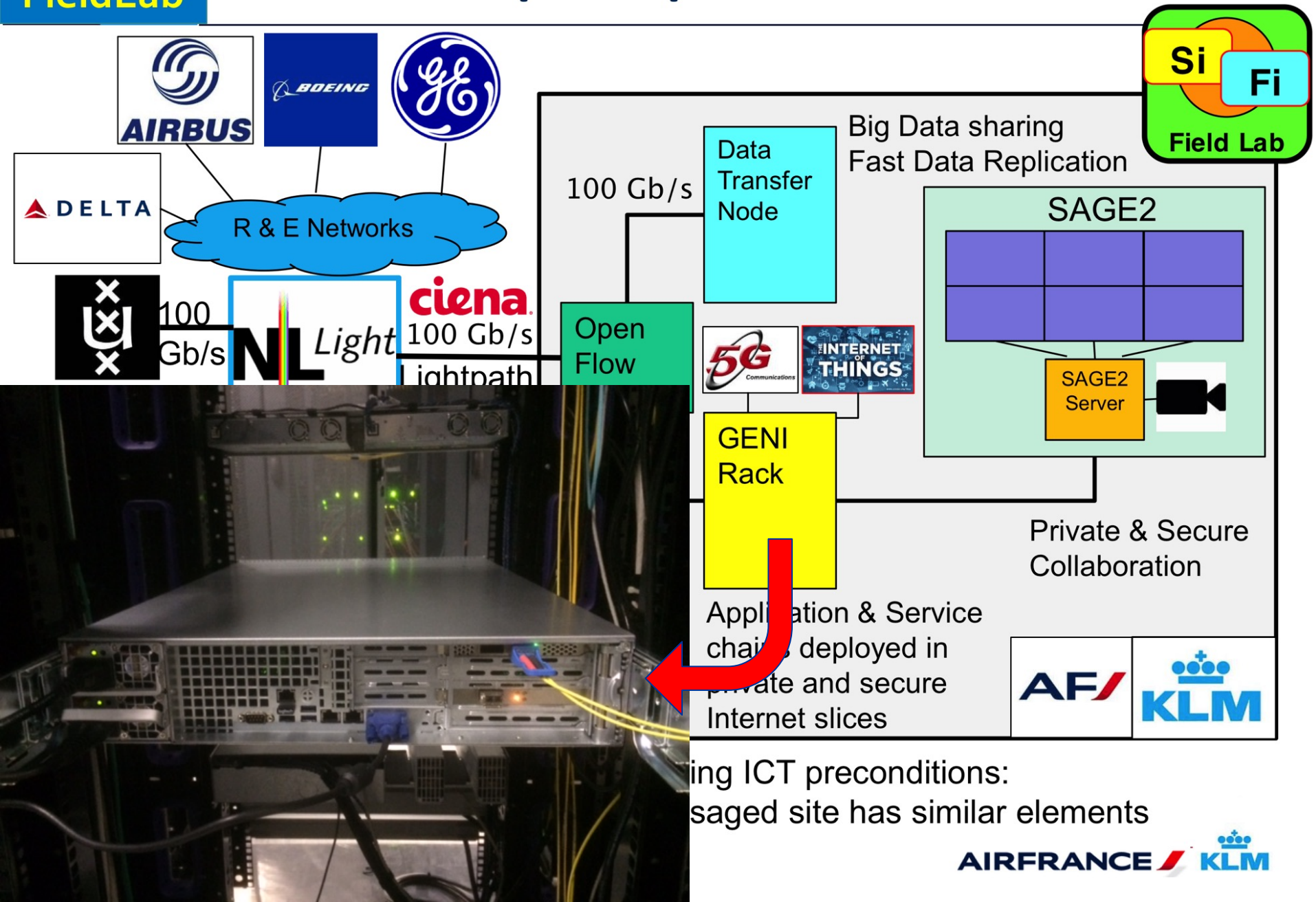


Re-enforcing ICT preconditions:
Each envisaged site has similar elements

AIRFRANCE KLM

AF/KLM FieldLab

Ambition to put capabilities into fieldlab

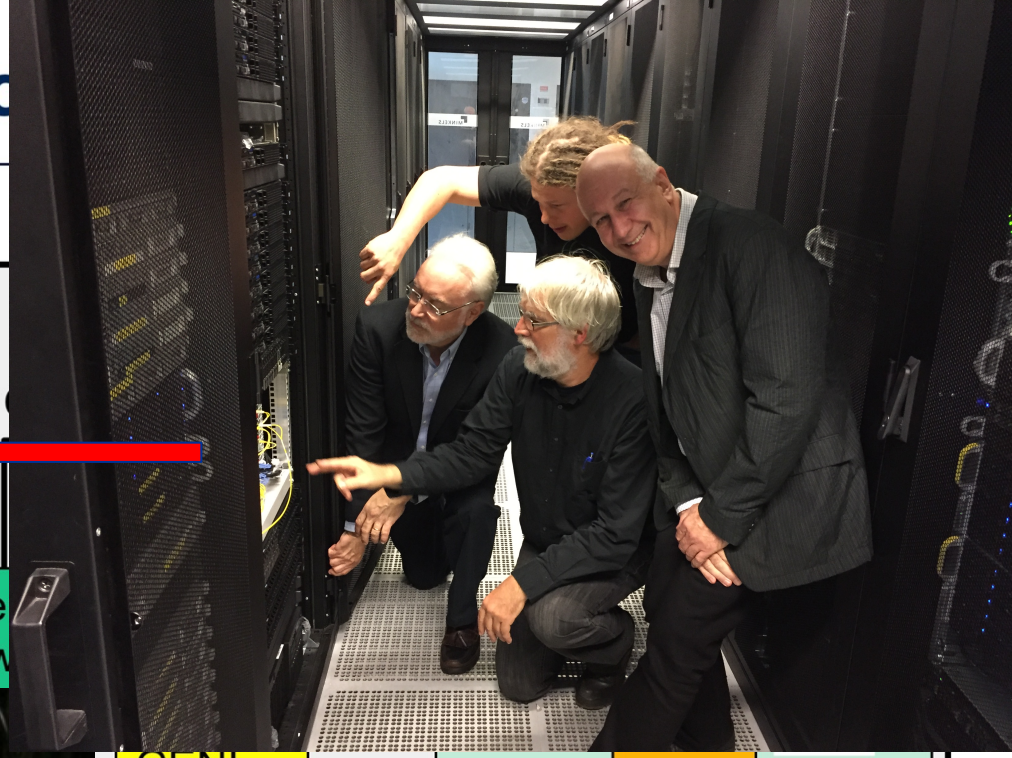


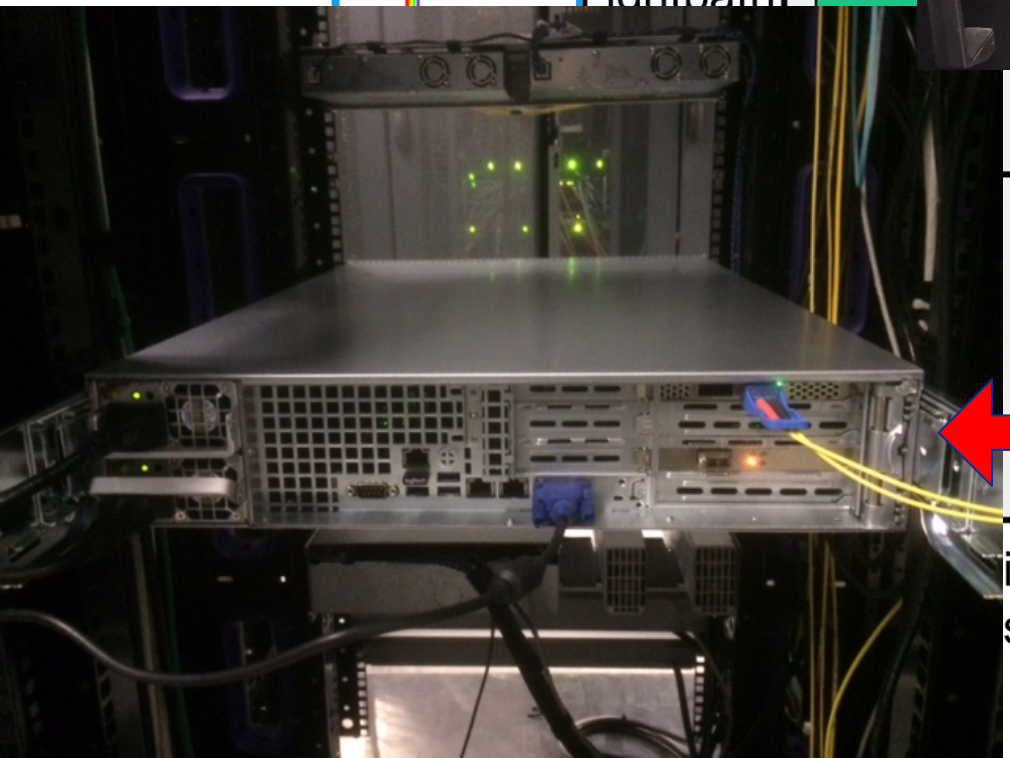
ing ICT preconditions:
saged site has similar elements

AIRFRANCE KLM

AF/KLM FieldLab

Ambition to put c





GENI Rack

Private & Secure Collaboration

Appli ation & Service chai deployed in vate and secure Internet slices

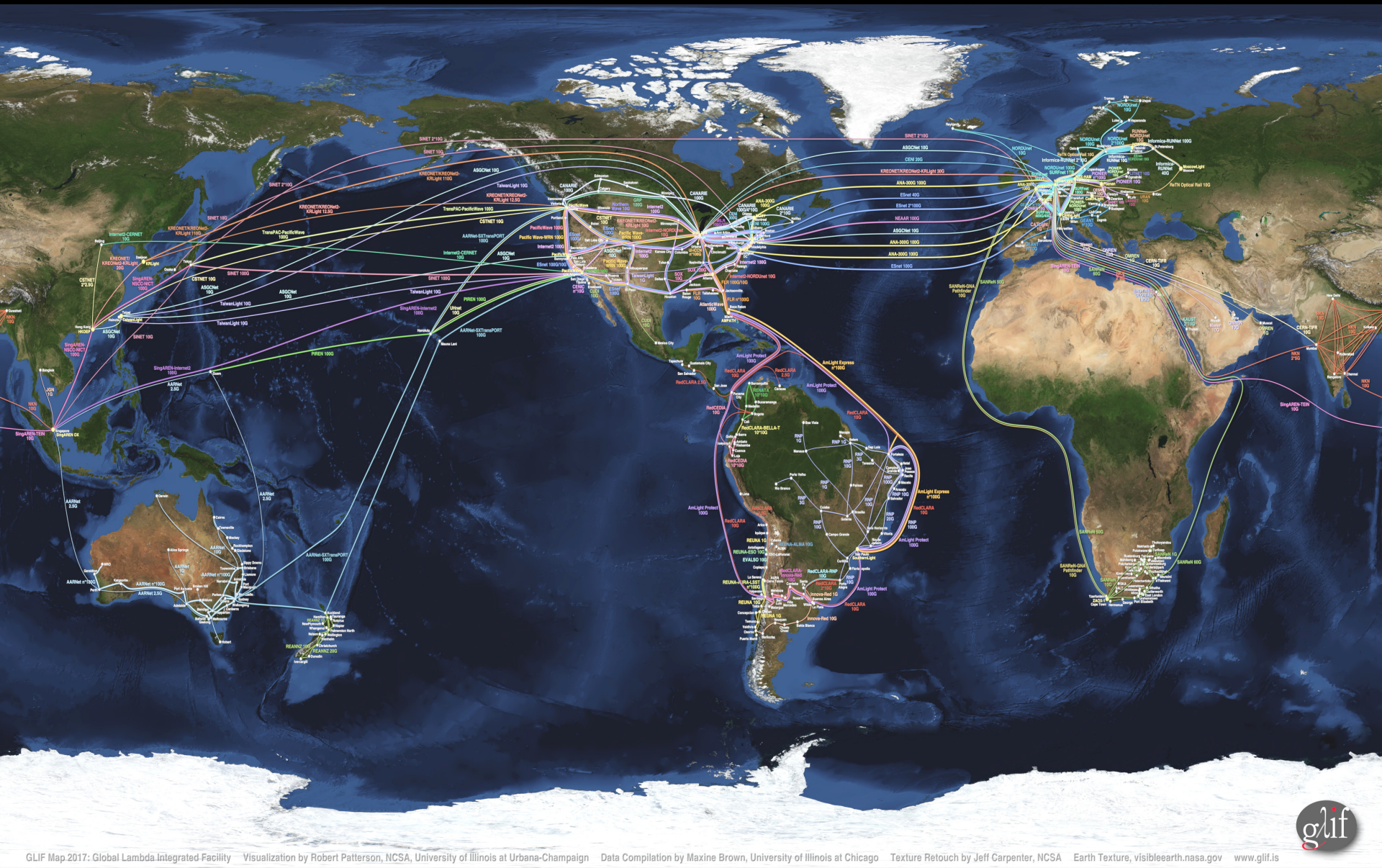
AF KLM

ing ICT preconditions:
saged site has similar elements

AIRFRANCE KLM

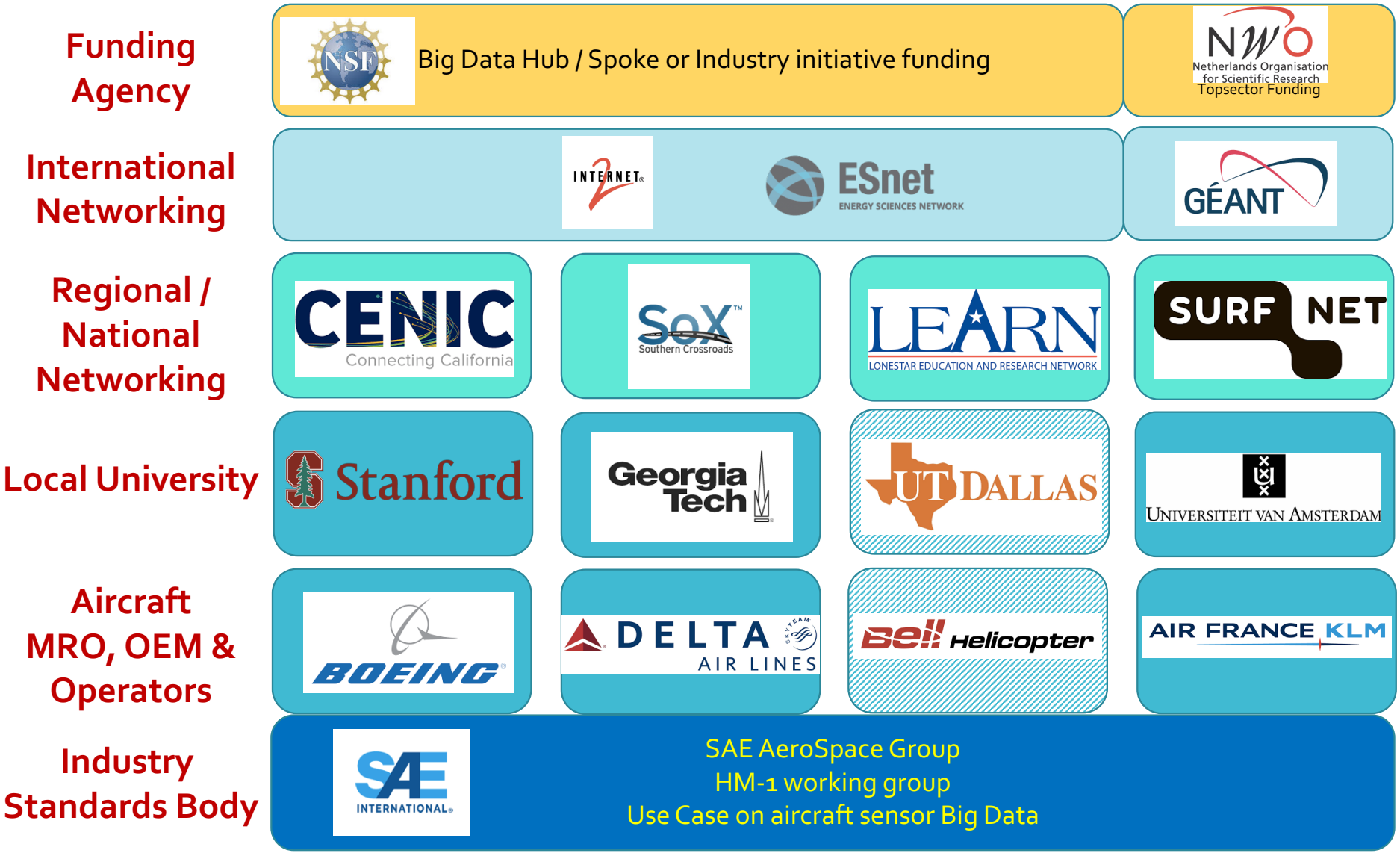
The GLIF – LightPaths around the World

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.



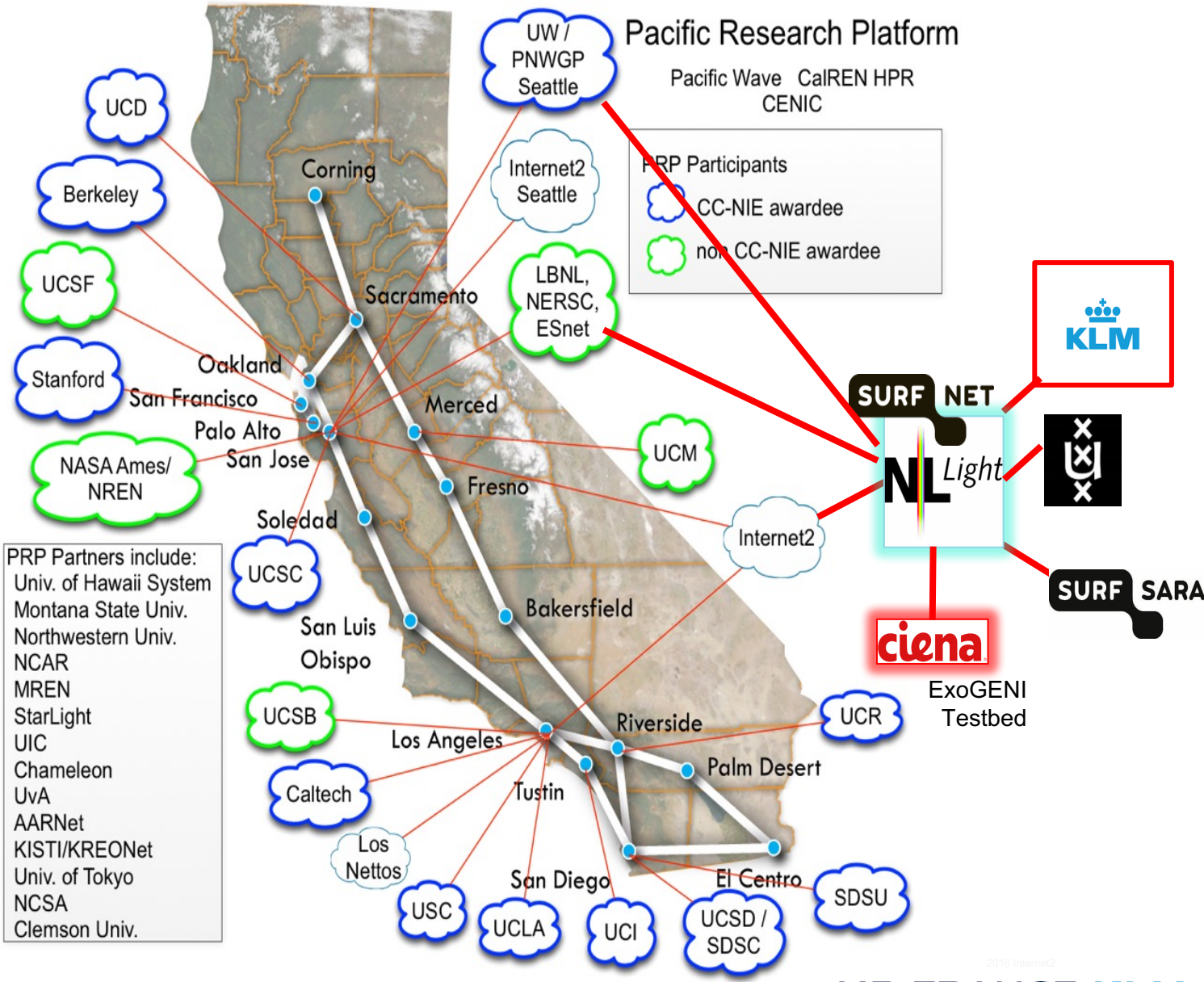
GLIF Map 2017: Global Lambda Integrated Facility Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign Data Compilation by Maxine Brown, University of Illinois at Chicago Texture Retouch by Jeff Carpenter, NCSA Earth Texture, visibleearth.nasa.gov www.glif.is

SAE Use Case envisaged research collaboration



Pacific Research Platform testbed involvement

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



PACIFIC RESEARCH PLATFORM

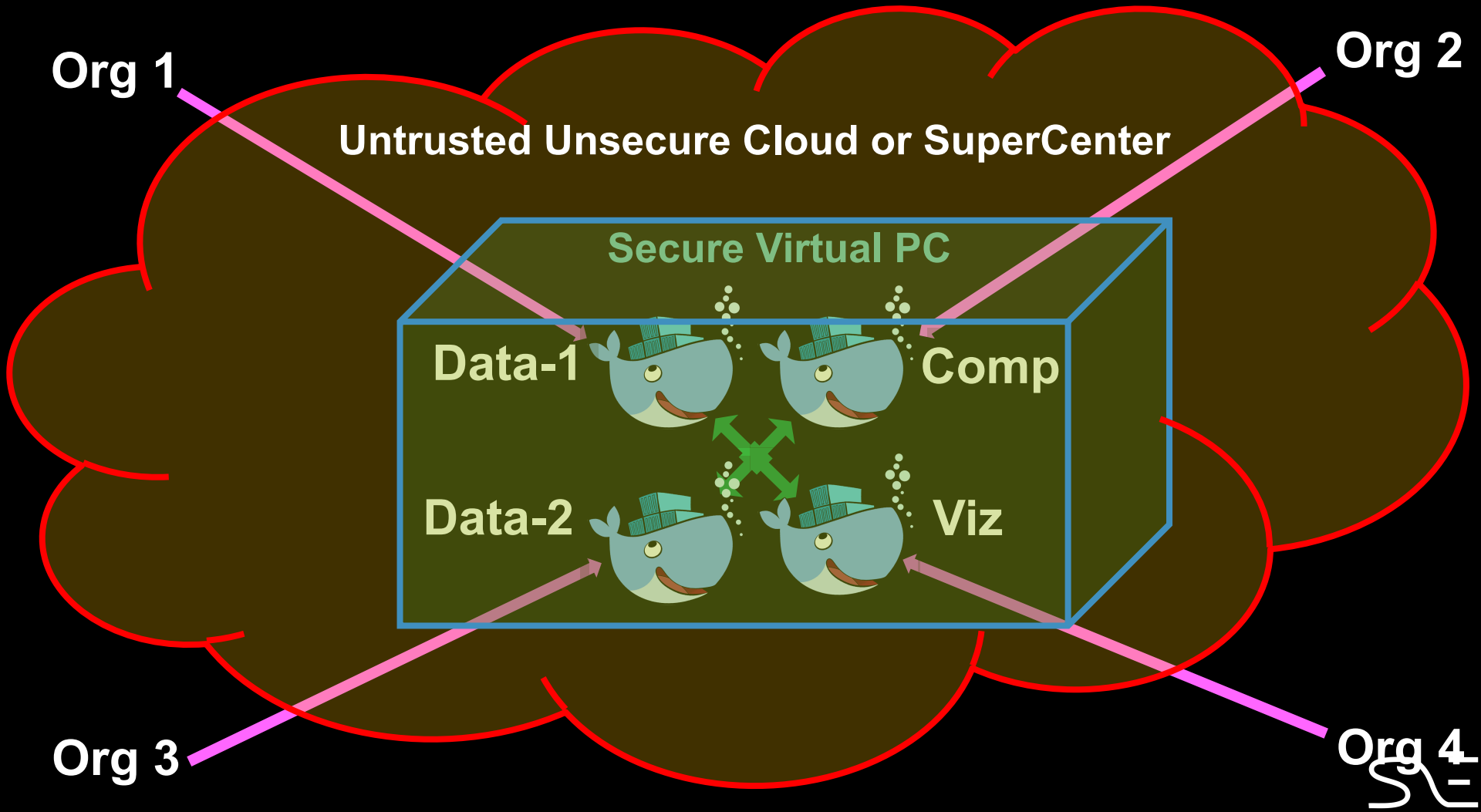
Note: this diagram represents a subset of sites and connections.

Approach

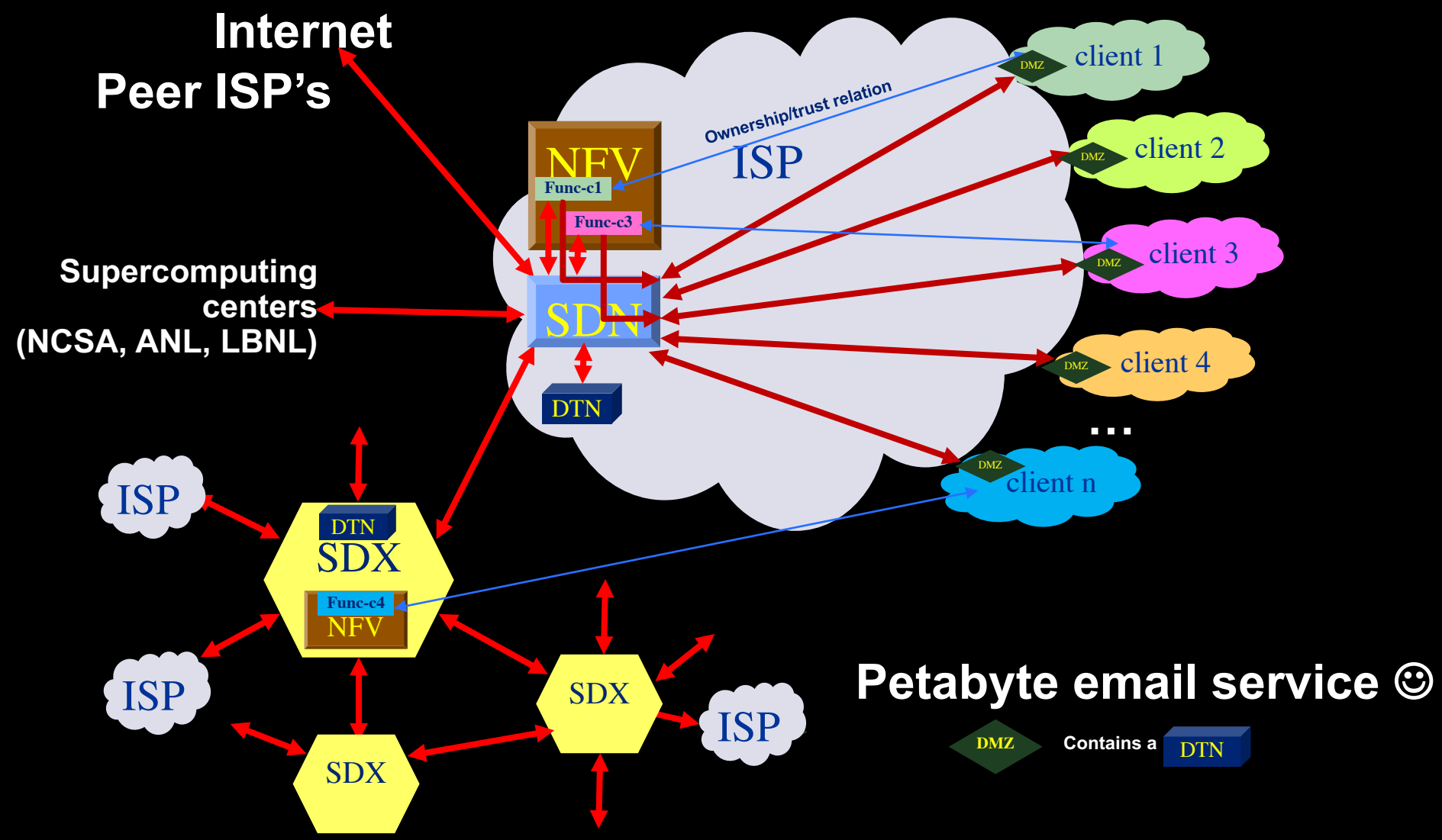
- Strategic:
  - Translate legislation into machine readable policy
  - Define data use policy
  - Trust evaluation models & metrics
- Tactical:
  - Map app given rules & policy & data and resources
  - Bring computing and data to (un)trusted third party
  - Resilience
- Operational:
  - TPM & Encryption 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 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 encryped hard drive)
- Compute Docker (protected application, signed algorithms)
- Visualization Docker (to visualize output)



Networks of ScienceDMZ's & SDX's



Q&A

- More information:
  - http://delaat.net/sarnet
  - http://delaat.net/dl4ld