

Global Collaborative Research Groups (CRGs)

The Dutch Big Data Hub infrastructure inspired by PRP
as model for European Open Science Cloud

@ NRP workshop, Bozeman (MT), 7 – 8 Aug, 2017

Cees de Laat

System & Network Engineering
University of Amsterdam



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

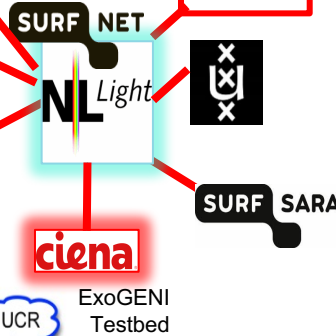
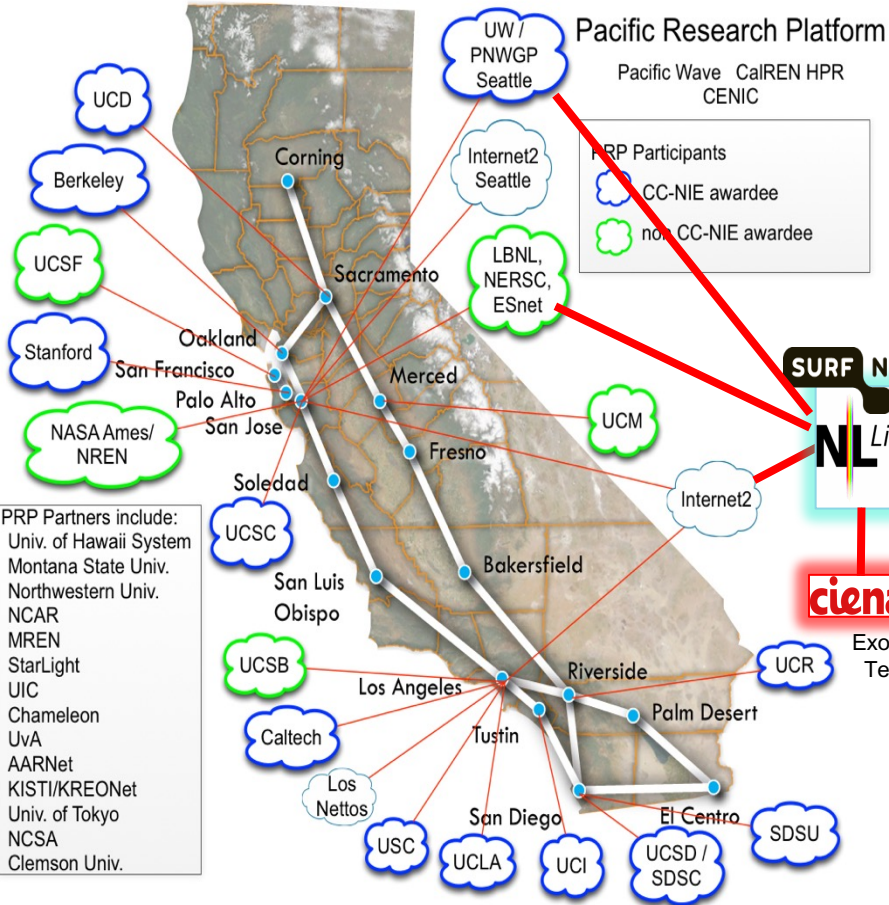


Pacific Research Platform testbed involvement

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



PRP Partners include:
 Univ. of Hawaii System
 Montana State Univ.
 Northwestern Univ.
 NCAR
 MREN
 StarLight
 UIC
 Chameleon
 UvA
 AARNet
 KISTI/KREONet
 Univ. of Tokyo
 NCSA
 Clemson Univ.



CdL Sabbaticals in spring 2015



1 - Berkeley - San Diego

All Photos sabb-1



2 - Washington - Chicago - Ottawa

All Photos sabb-2



Activity

work

leisure

Purpose

This personal website contains logs of my sabbatical activities in the spring of 2015. I went on these sabbaticals to be able to engage the community based on content in stead the more and more managerial stuff I was sucked into at UvA. Also my participation in the SURF scientific advisory board made it necessary to study future directions of cyber infrastructure. Apart from those I always keep on my toes if the research questions we as a group at UvA are studying, are still novel and valid. Directions change very rapidly particular field of study in System and Network Engineering.

Goals

- Discuss with peers the 5 and 10 year outlook for Cyber Infrastructure. Budgets are shrinking, networks become mature, Supercomputing and High Throughput data compressing is now also done with huge commercial cloud centers, so what is the position of specific mission resources in this wild ocean of public capabilities. What do science community need to do ourselves and what can we just buy in the future from the (cloud) market. What do and what do we not need to do on leadership supercomputing. How do we relate to public cloud. We also do not have scientific water. What do we need to do on data at the central level? Do we need a national data service including data stewardship, etc.
- Next year we are chairing the eIRG workshop. The themes will be CI directions and funding models on national and european level. During the sabbatical seeking out potential speakers.
- Reflection on our own research questions

Personal

- Take some holiday
- Have Emelie joining me for a couple of days
- Drive the HWY1
- Visit the desert
- Go to Chicago theatre, music bars, etc.

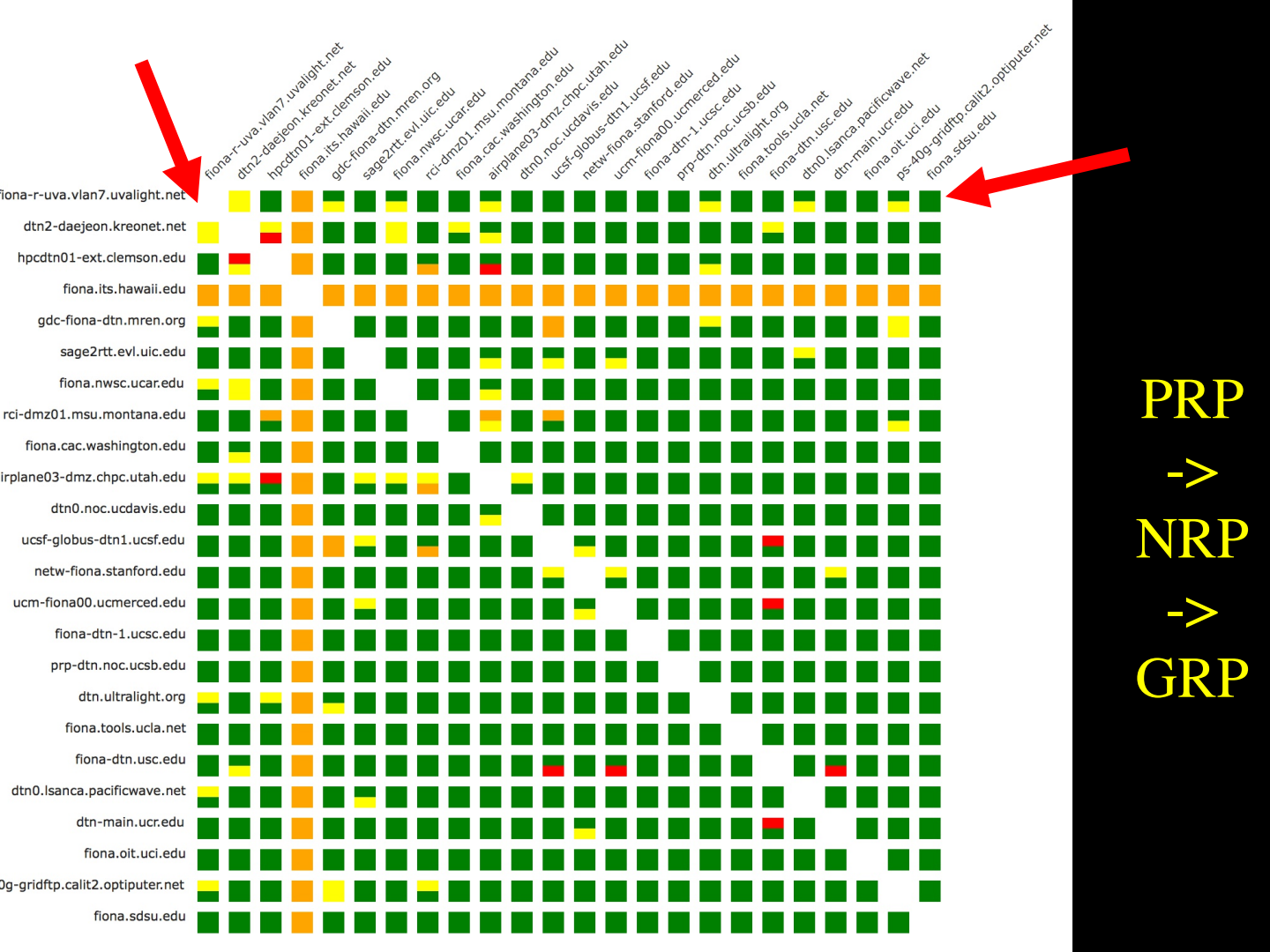


Why me?



Was there on sabbatical





PRP

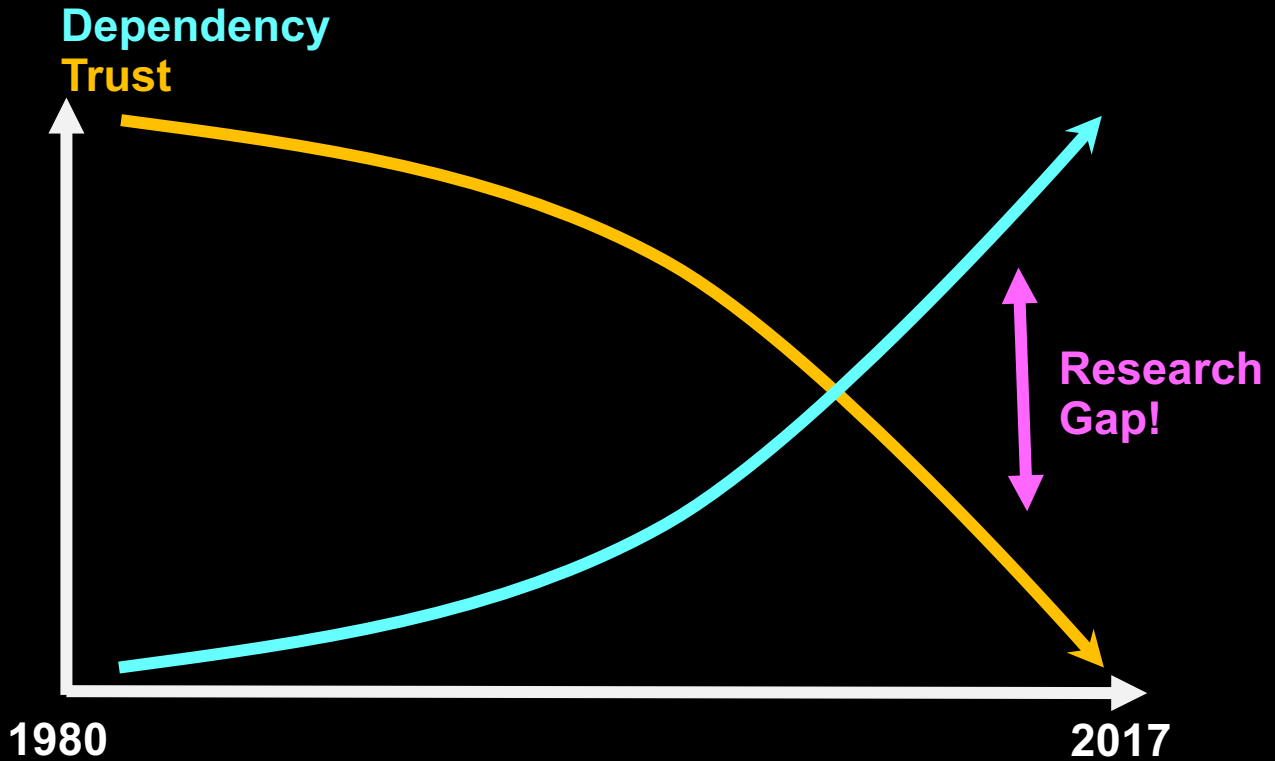
->

NRP

->

GRP

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



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

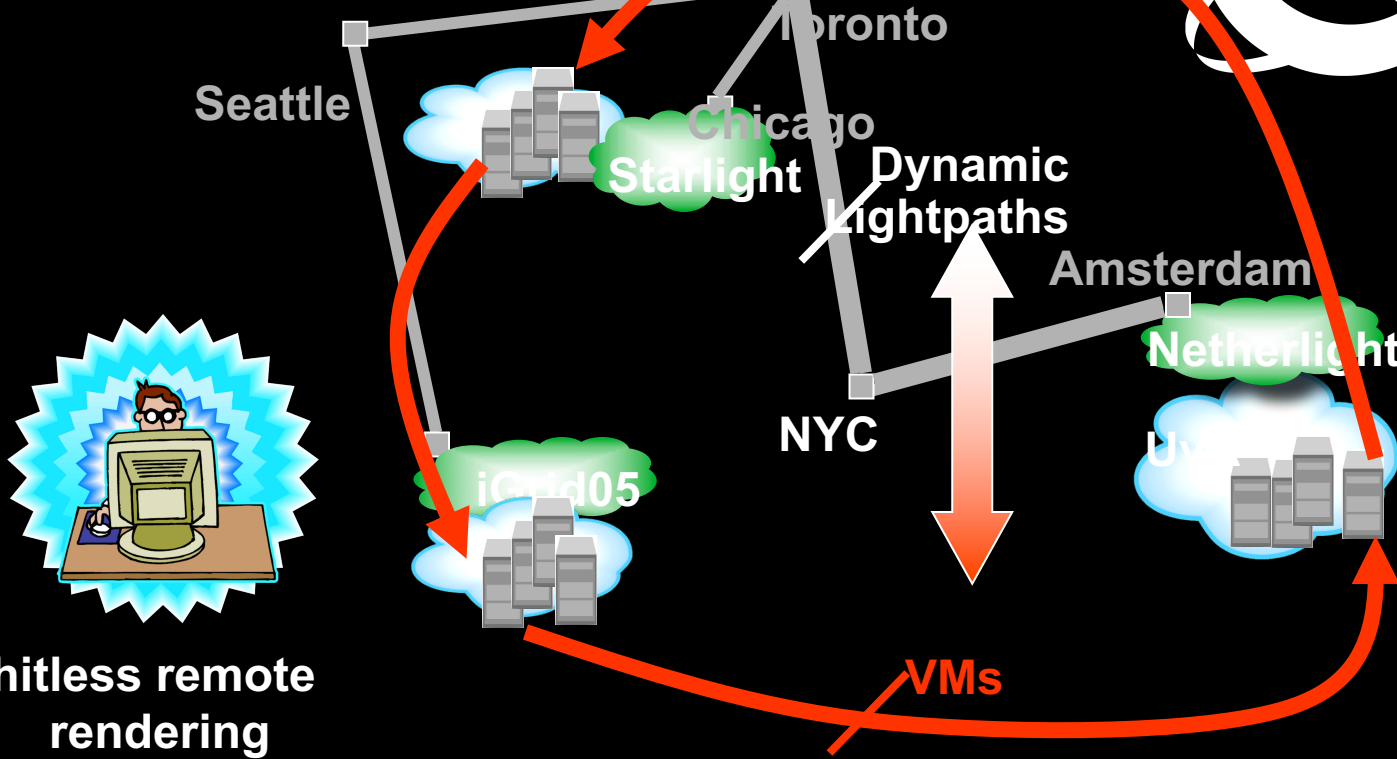


SAE AeroSpace Group
HM-1 working group
Use Case on aircraft sensor Big Data

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

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

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

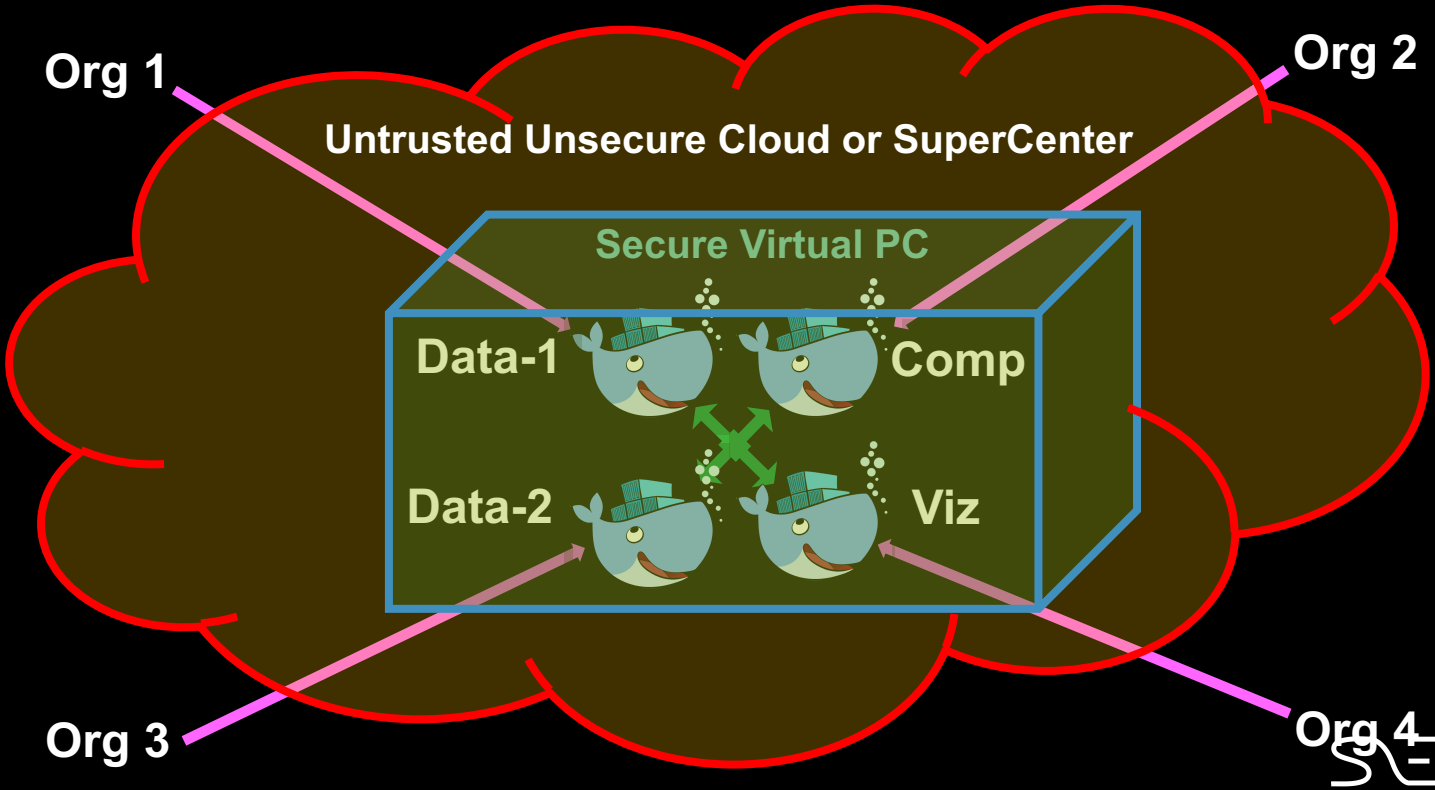
Comp

Data-2

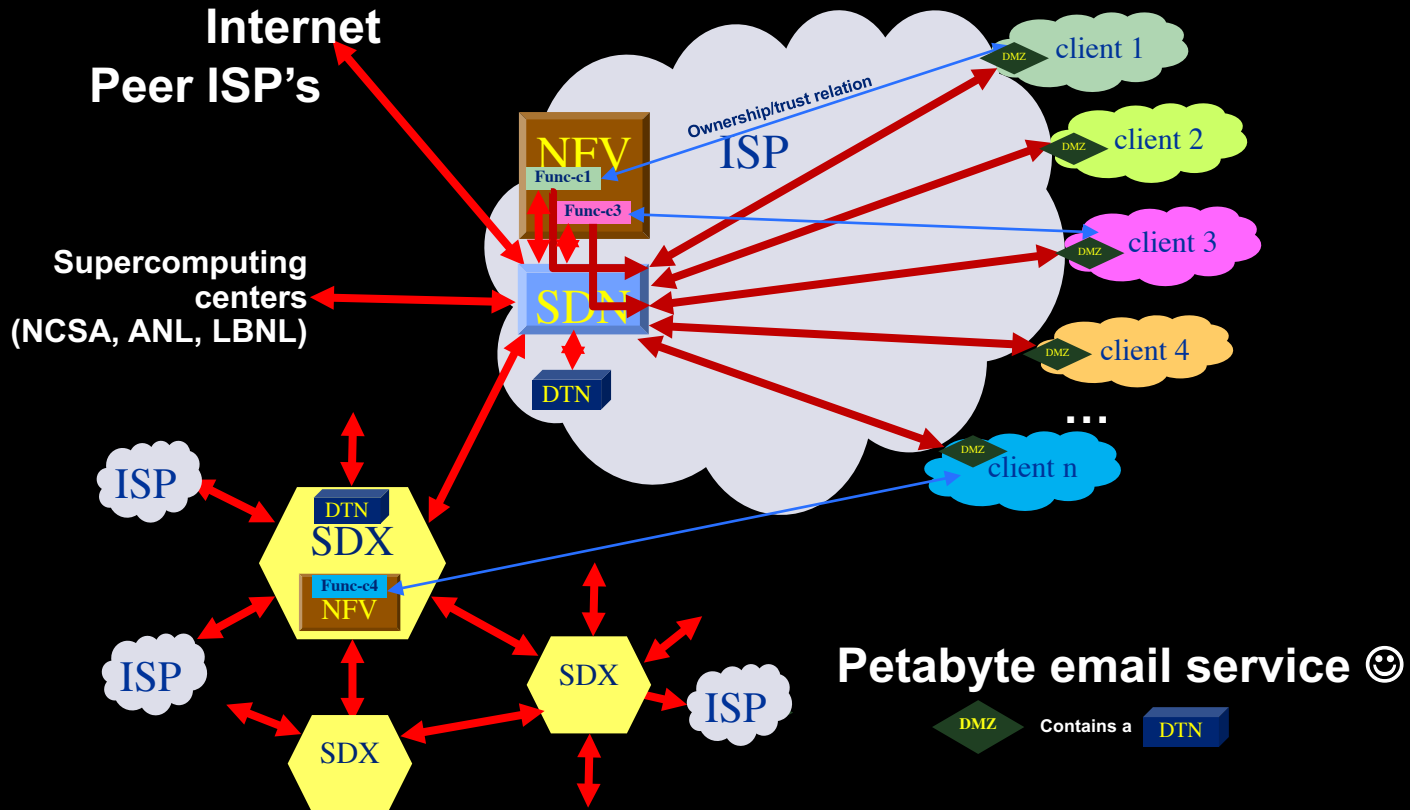
Viz

Org 3

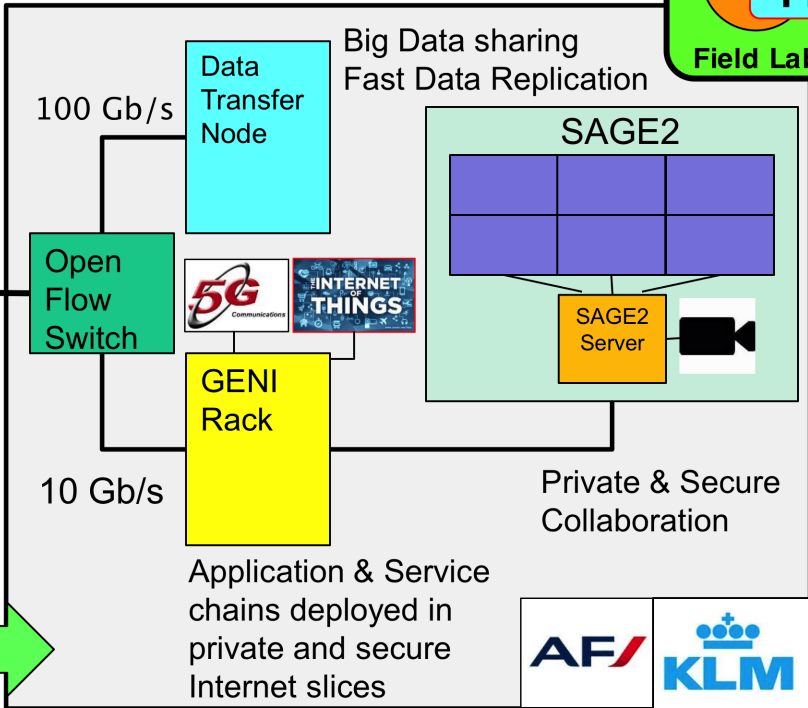
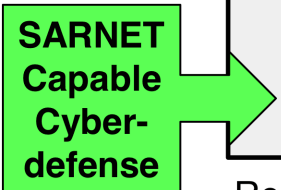
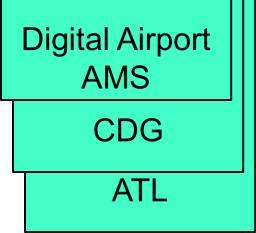
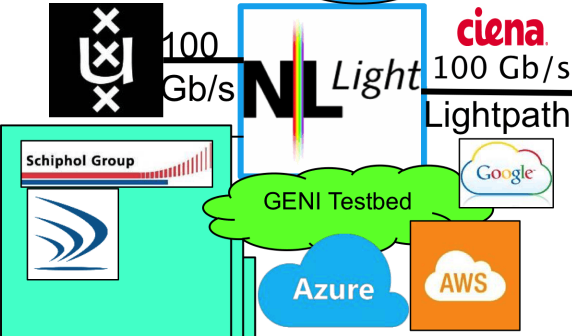
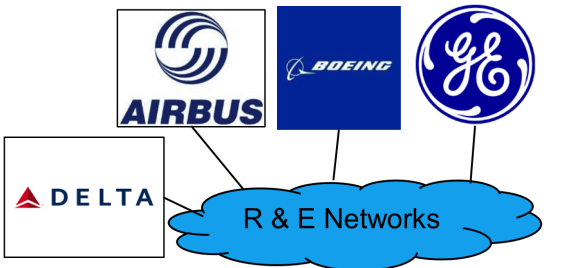
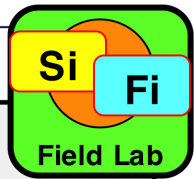
Org 4



Networks of ScienceDMZ's & SDX's



Ambition to put capabilities into fieldlab



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



Data Hub System Applicability

Industry

- Cross Cutting Field lab
- Innovation with SURF



Science

- European Open Science Cloud
- FAIR model



Society

- Smart Cities & Arena
- Streaming Data Decision Support



Validation Fieldlab and Dissemination

UVA - OpenLab

KLM
NetherLight
GENI
Fed4Fire
Cloud
SURFSARA
...



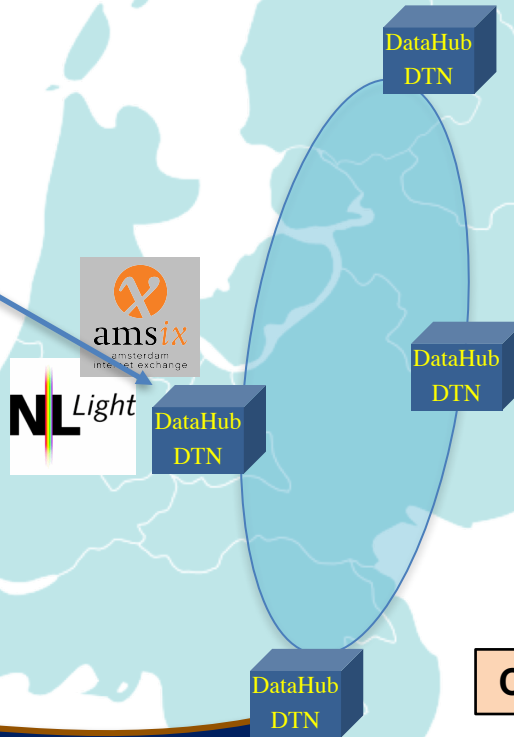
TNO - Intrepid

Smart Data
Factory
Innovations
Smart Rail
To-Grip
...

C2D – Big DataHubs

Arena
KAVE
AZURE
Use Cases
...

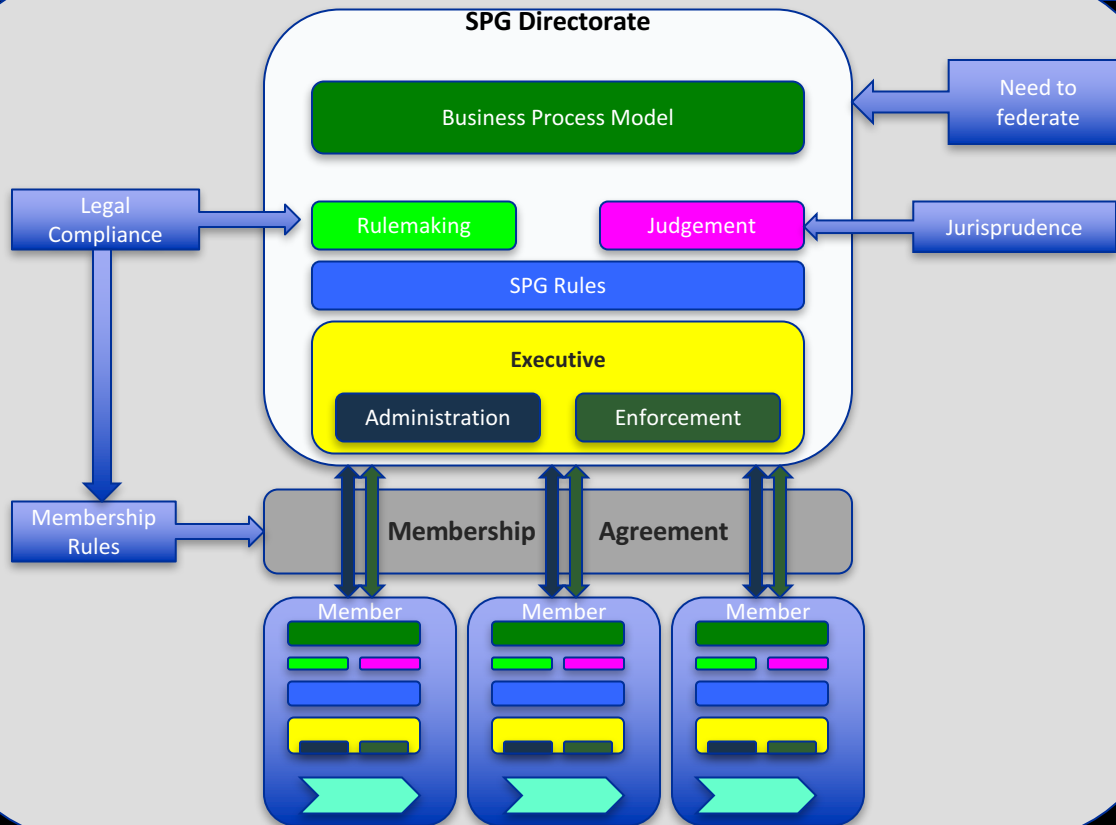
- Experimental facilities from day one!
- Proof of concepts demonstrating secure data sharing
- Blueprint, roadmap and standards where applicable
- Model for FAIR EOSC Infrastructure



International research collaborations

- Raise extra challenges:
 - The long RTT and the higher amount of interconnection networks make network performance a never ending problem.
 - E.g. we had faulty photonics and somewhere in one direction in the network a problem.
 - Ensure all inside a big collaboration technologically work together?
 - How to organize security and trust inside and amongst such collaborations?
 - Different time zones.
- Ultimately:
 - How to have different collaborations work together for system level science?
 - E.g. PIRE - Open Science Data Cloud
 - European Open Science Cloud
 - → FAIR
 - → Findable – Accessible – Interpretable - Reusable

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

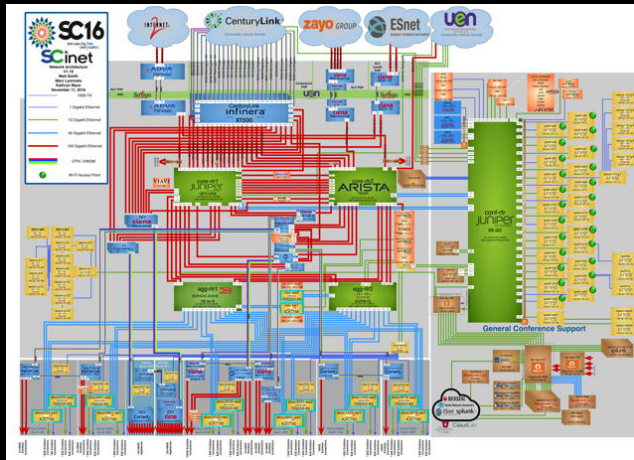
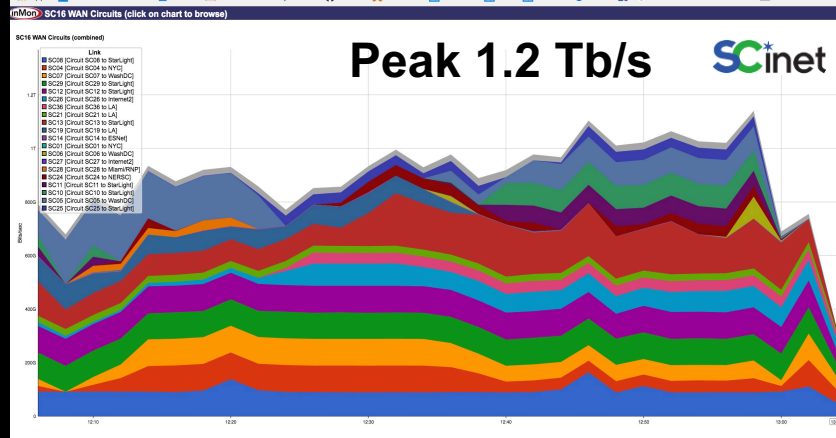


Pushing PRP & Science DMZ models

- UvA writing Campus CI plan, has ScienceDMZ established
 - ScienceDriver: genomics
- UvA, KLM, SURFnet have DTN's
- SURFnet challenging all universities to follow the lead.
- Dutch NSF and eScience center asks CI requirements in proposals. Advocating this also in the EU.
- Benefits:
 - Requires CIO's to think about supporting science and not only mostly education and management.
 - Gives the NREN and Science Foundation an idea about where in the country are the hotspots in CI and requirements. -> feeds design!
 - Displays the ambitions.

Learned from Scinet & INDIS

- 2013 - 2016
 - SDN
 - Security
 - Traffic management, policing, control
 - Hybrid – optical ring - approach to reach Tb/s
- 2017 - 2020
 - NFV
 - SDX
 - DTN @ core → petabyte email network
 - Data abstractions (e.g. NDN)



INDIS

Innovating the Network for Data-Intensive Science



SCinet

[Home](#)

[About](#)

[Papers](#)

[Program](#)

[Presentations](#)

[Important Dates](#)

[Previous INDIS Workshops](#)

Papers

INDIS workshop call for papers now open! Deadline September 5

[Read more](#)

SC17 INDIS Workshop

Posted by scinet-sysadmin on Wednesday, 7 June 2017

The 4th International Workshop on Innovating the Network for Data Intensive Science (INDIS) 2017 will be held in conjunction with SC17: IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis, in Denver, CO, on November 12, 2017, 9h00 – 17h30.

[Read more](#)

<https://scinet.supercomputing.org/workshop/>

Quick Links

- [Important Dates](#)
- [Program](#)
- [Previous INDIS Workshops](#)

User login

Username *

Password *

- [Request new password](#)

[Log in](#)