ONDM 2009:
Topology handling in GLIF

Cees de Laat

GLIF.is founding member

SURFnet

EU

BSIK

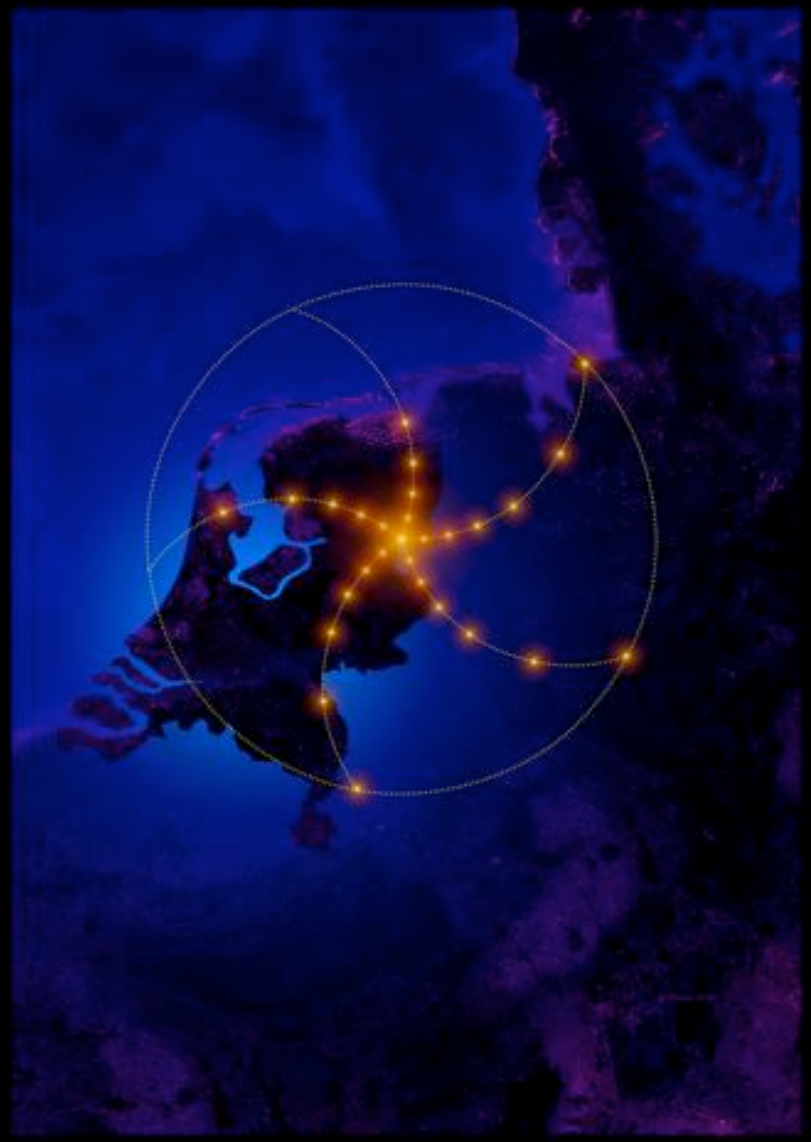
NWO

University of Amsterdam

TNO

NCF

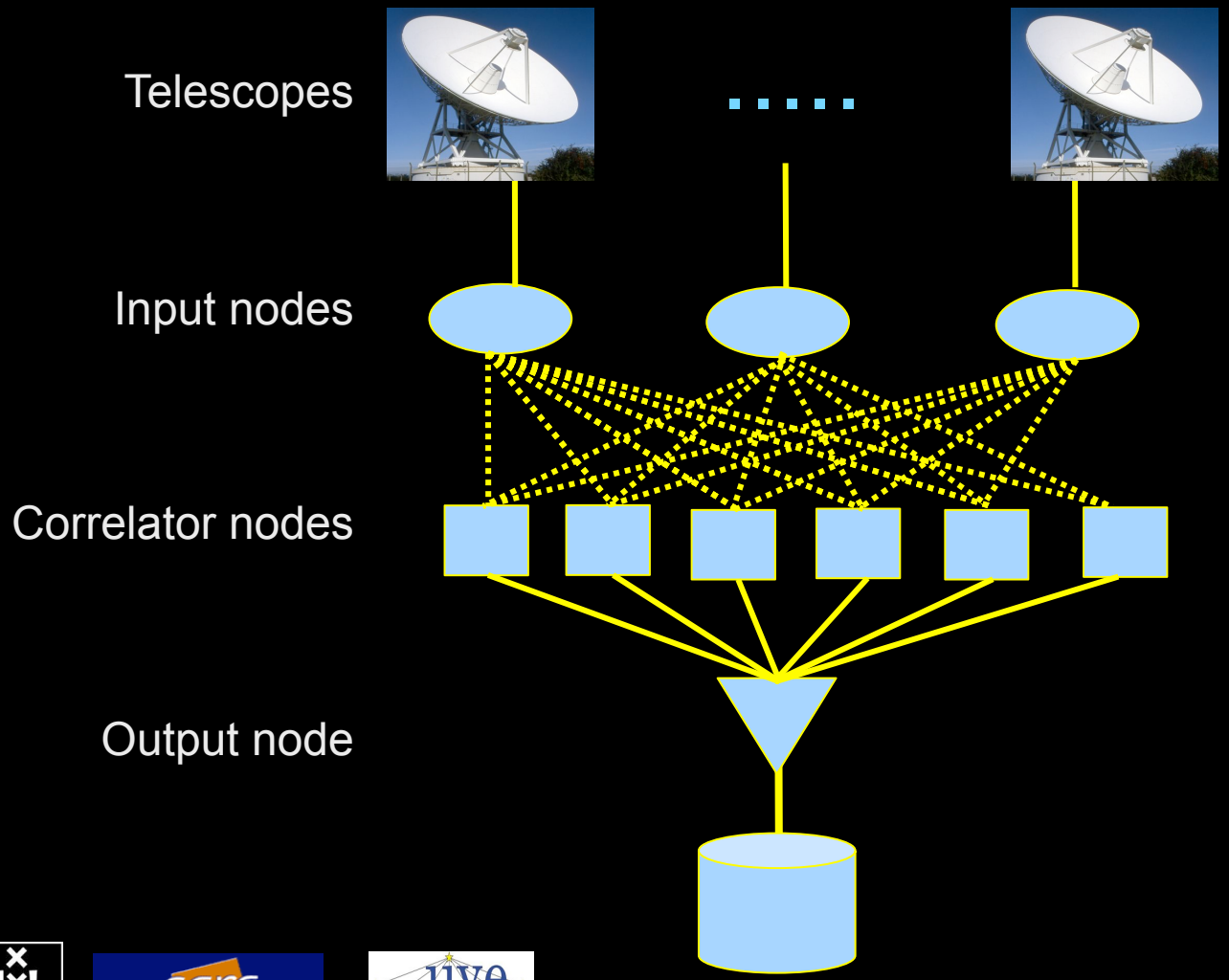
LOFAR as a Sensor Network



- LOFAR is a large distributed research infrastructure:
  - Astronomy:
    - >100 phased array stations
    - Combined in aperture synthesis array
    - 13,000 small “LF” antennas
    - 13,000 small “HF” tiles
  - Geophysics:
    - 18 vibration sensors per station
    - Infrasound detector per station
  - >20 Tbit/s generated digitally
  - >40 Tflop/s supercomputer
  - innovative software systems
    - new calibration approaches
    - full distributed control
    - VO and Grid integration
    - datamining and visualisation

The SCARIe project

SCARIe: a research project to create a Software Correlator for e-VLBI.
VLBI Correlation: signal processing technique to get high precision image from spatially distributed radio-telescope.



To equal the hardware correlator we need:

16 streams of 1Gbps

16 * 1Gbps of data

2 Tflops CPU power

2 TFlop / 16 Gbps =

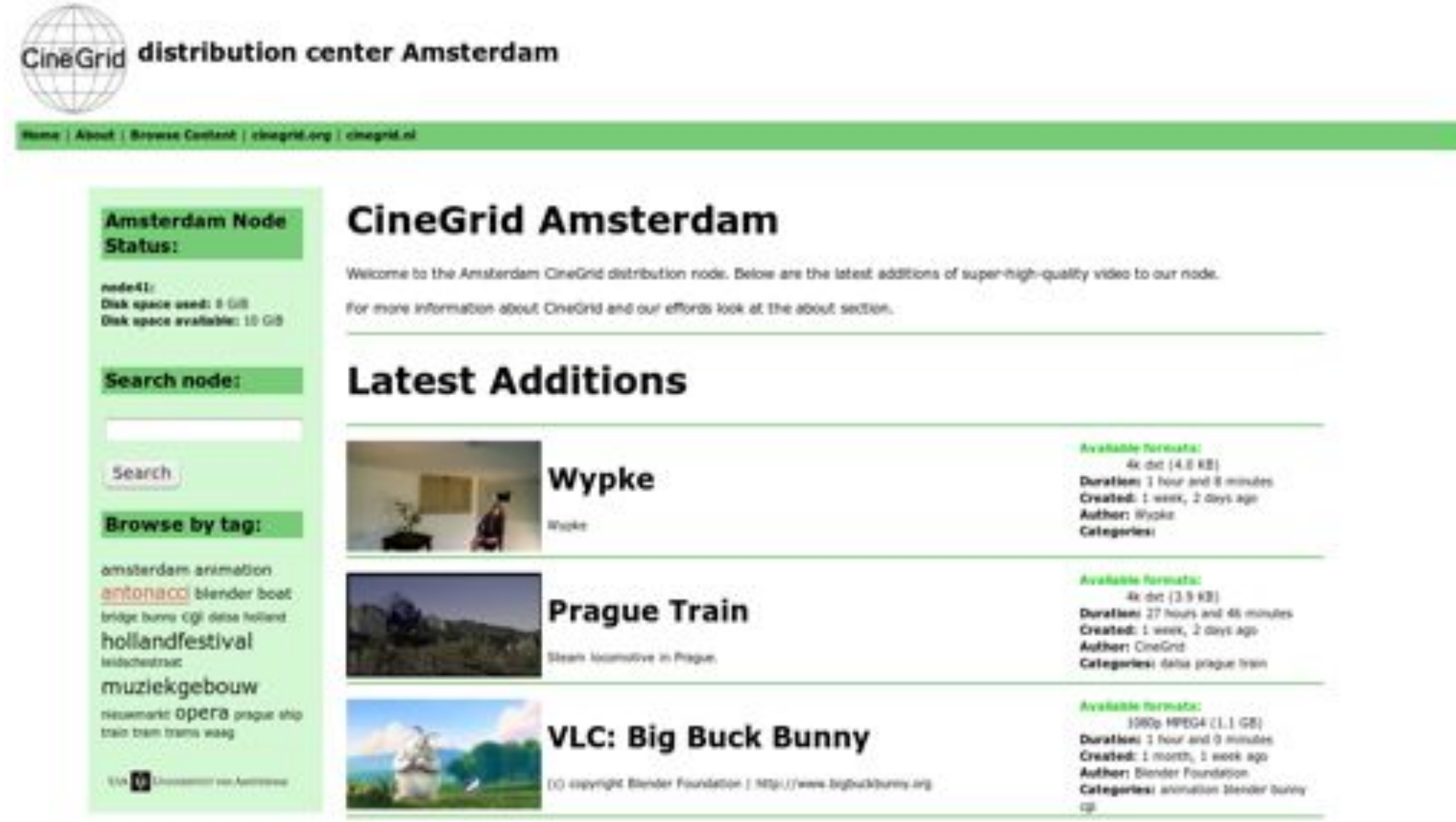
1000 flops/byte

THIS IS A DATA FLOW PROBLEM !!!

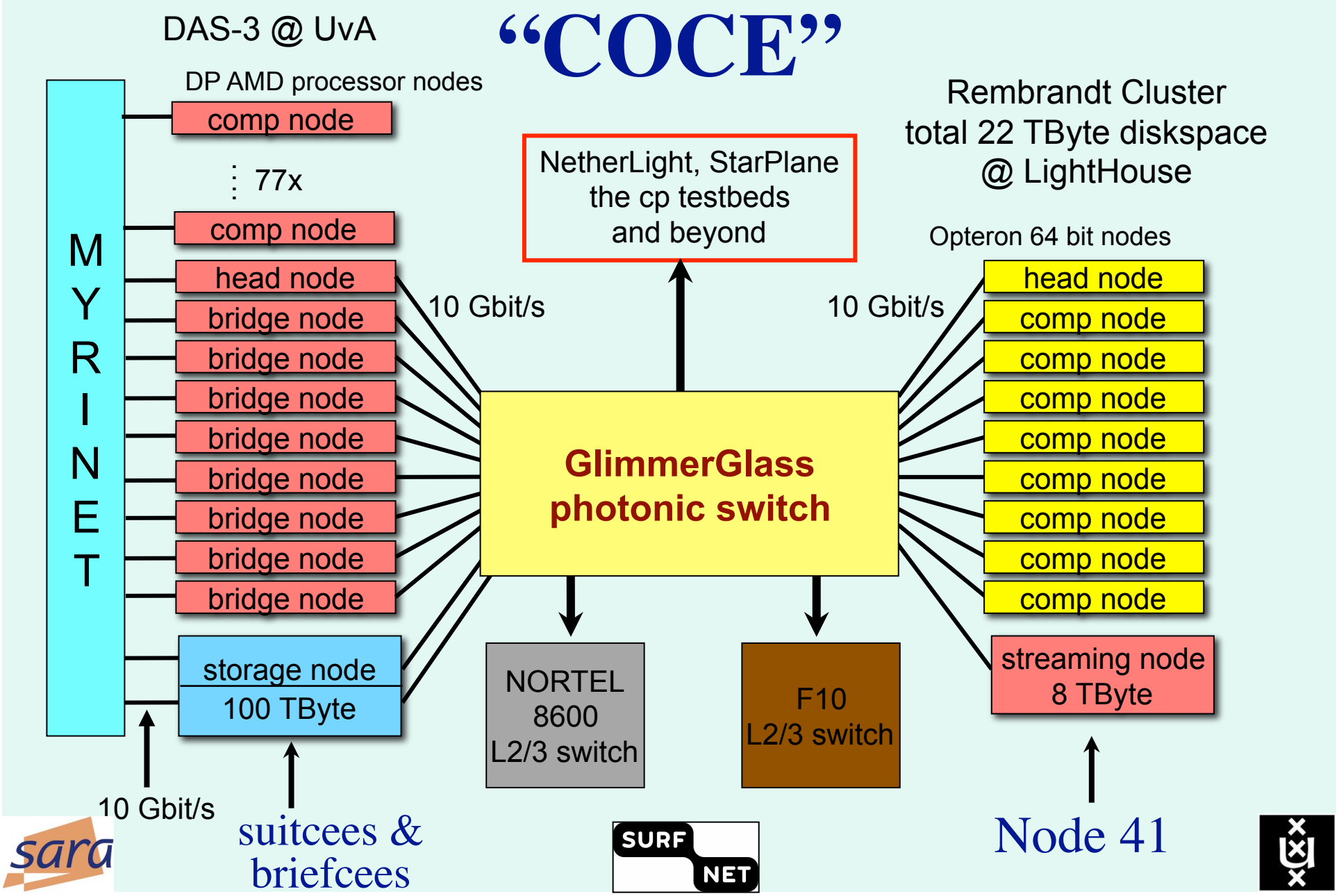
sara Computing & Networking Services

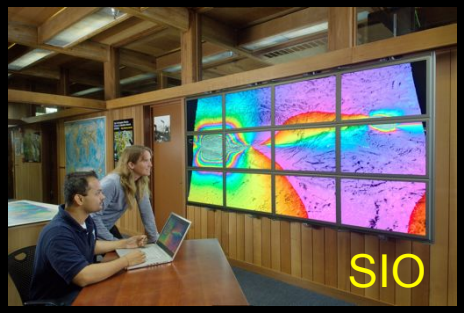
jive JOINT INSTITUTE FOR VLBI IN EUROPE

CineGrid portal



Amsterdam CineGrid S/F node "COCE"




SIO

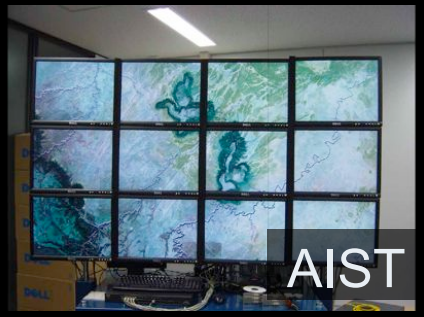

NCMIR


GeoWall 2
USGS EDC

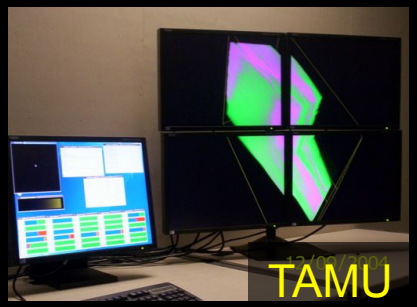

NCSA &
TRECC


SARA


KISTI


AIST


RINCON & Nortel


TAMU


UCI


UIC


CALIT2

IJKDIJK

Sensor grid: instrument the dikes

First controlled breach occurred on sept 27th ‘08:

30000 sensors (microphones) to cover all Dutch dikes

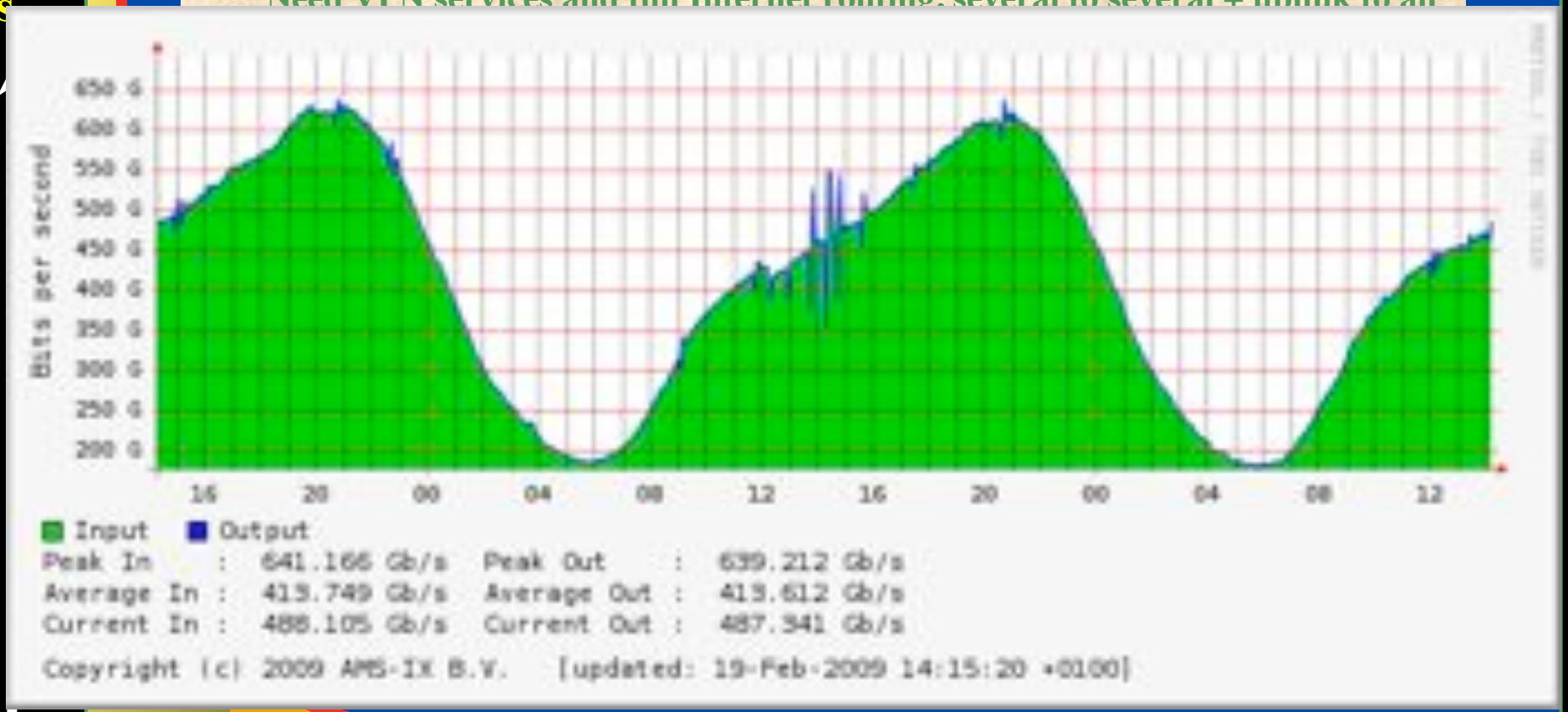
#users

A. Lightweight users, browsing, mailing, home use

Need full Internet routing, one to all

B. Business/grid applications, multicast, streaming, VO's, mostly LAN

Need VPN services and full Internet routing, several to several + uplink to all



B

C

ADSL (12 Mbit/s)

GigE

BW requirements

Towards Hybrid Networking!

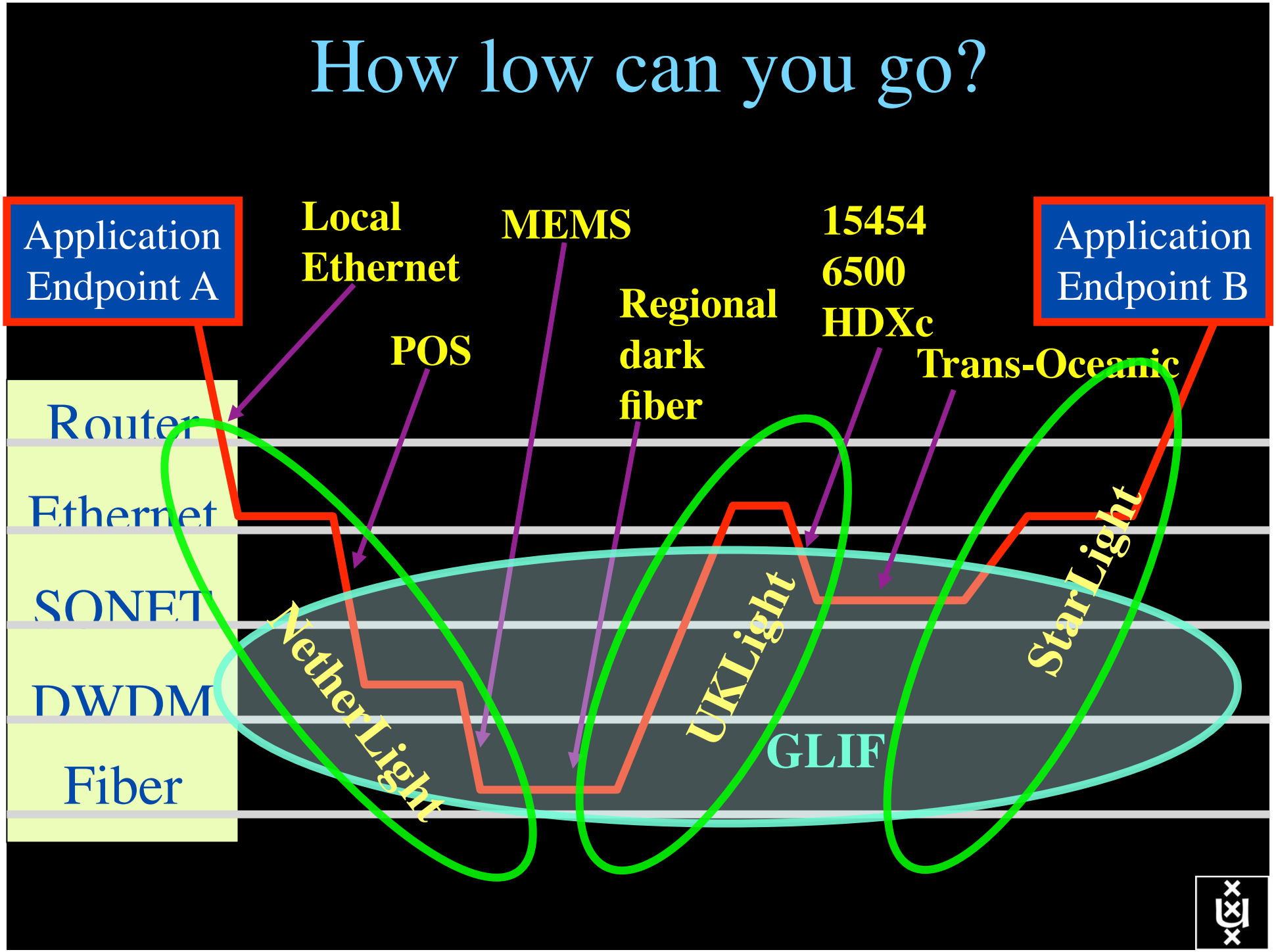
- Costs of photonic equipment 10% of switching 10 % of full routing
  - for same throughput!
  - Photonic vs Optical (optical used for SONET, etc, 10-50 k$/port)
  - DWDM lasers for long reach expensive, 10-50 k$
- Bottom line: look for a hybrid architecture which serves all classes in a cost effective way
  - map A -> L3 , B -> L2 , C -> L1 and L2
- Give each packet in the network the service it needs, but no more !

L1 ≈ 2-3 k$/port
0.5 W/port

L2 ≈ 5-8 k$/port
10-15 W/port

L3 ≈ 75+ k$/port
250 W/port

How low can you go?



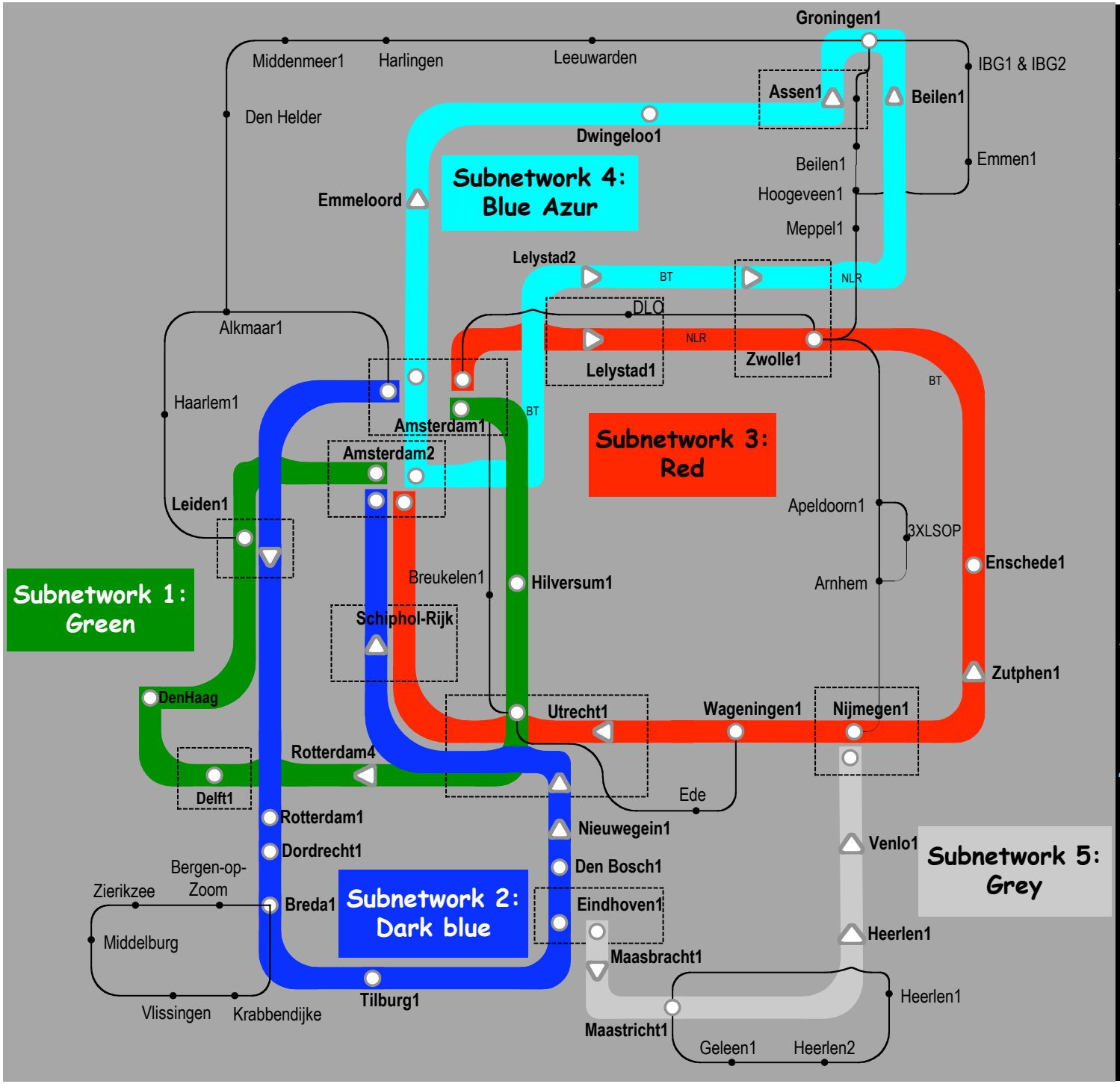
In The Netherlands SURFnet connects between 180:

- universities;
- academic hospitals;
- most polytechnics;
- research centers.

with an indirect ~750K user base

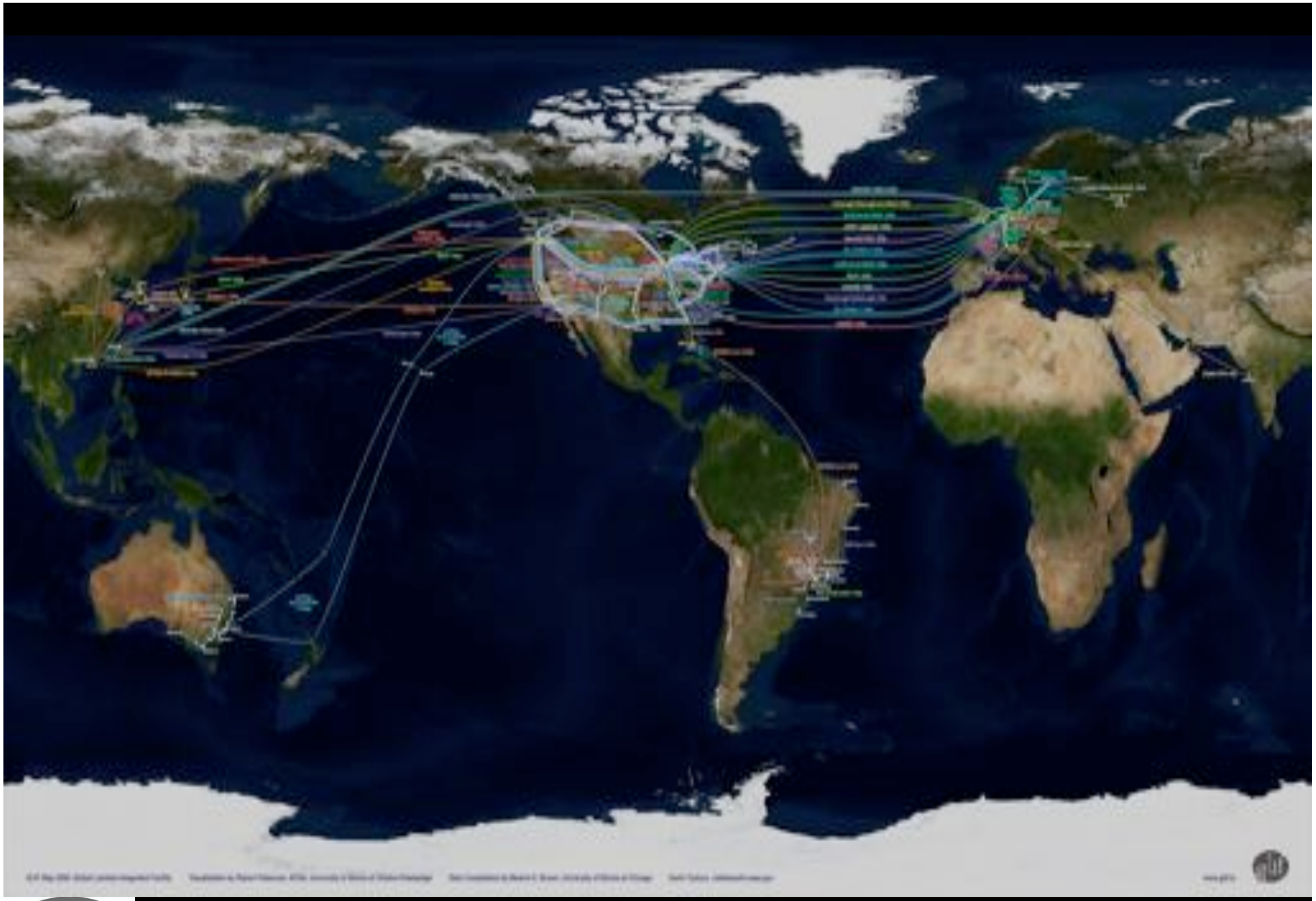
~ 8860 km

scale comparable to railway system



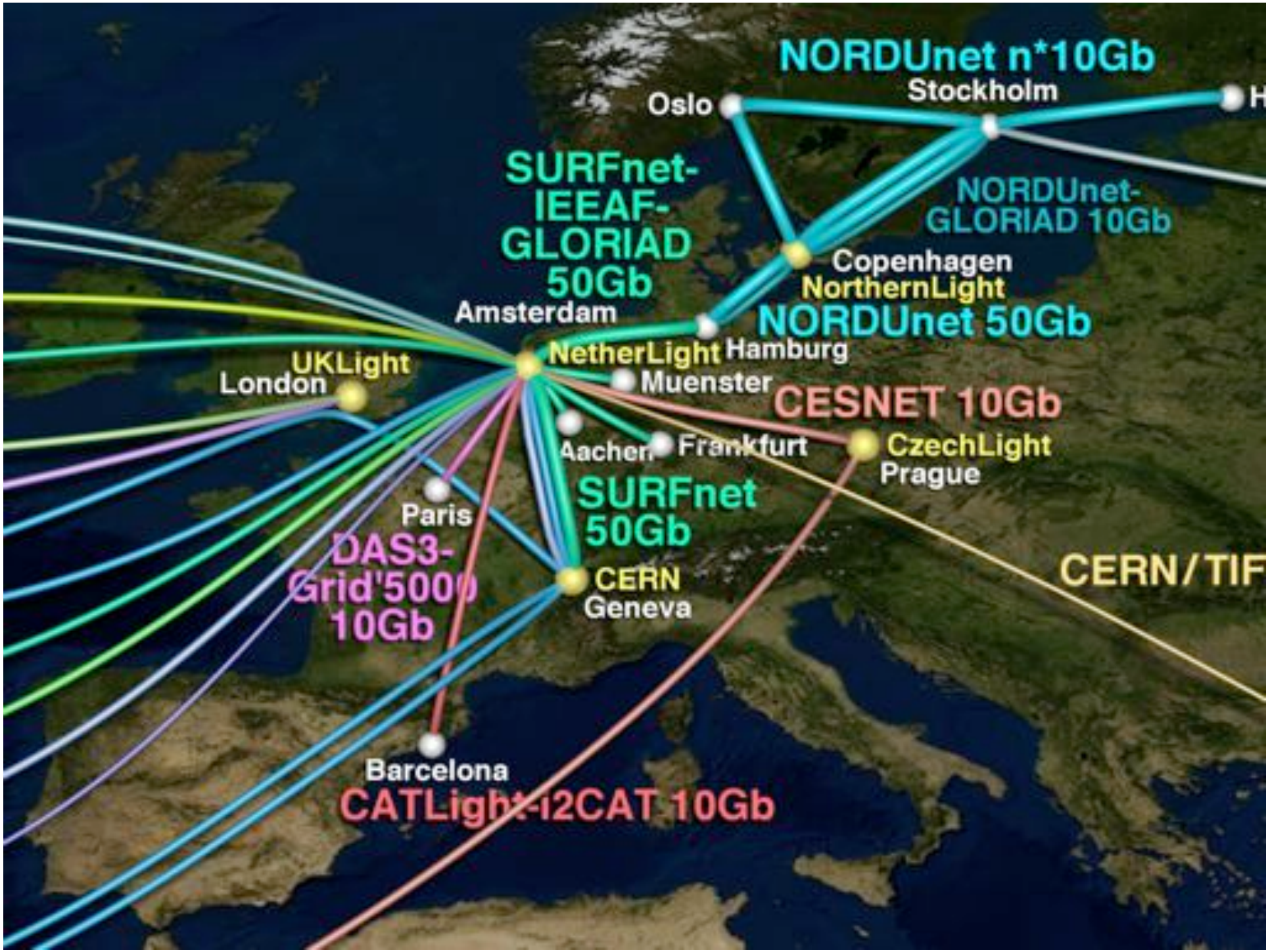
Common Photonic Layer (CPL) in SURFnet6

supports up to
72 Lambda's of
10 G each
40 G soon.



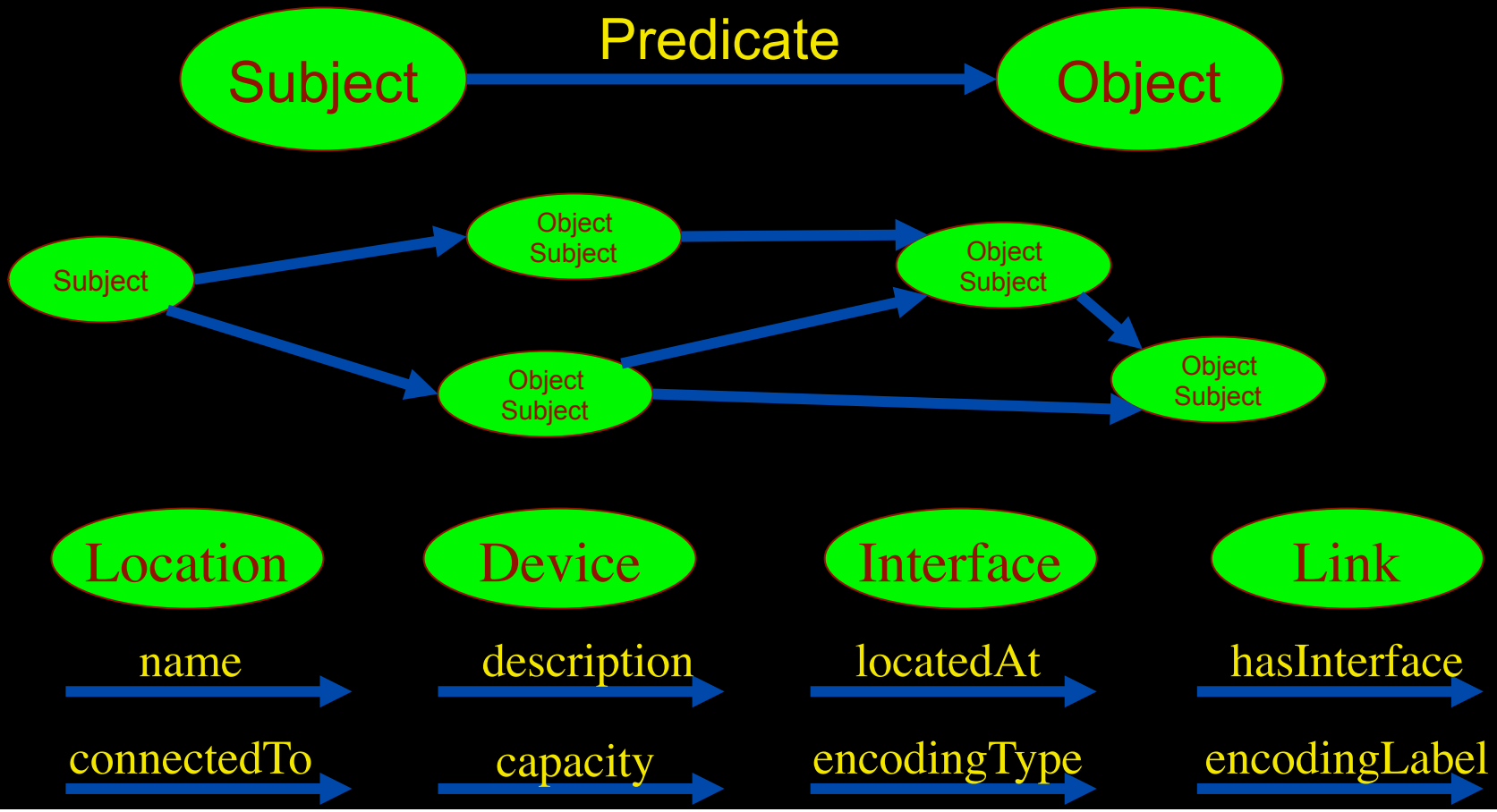
GLIF 2008

Visualization courtesy of Bob Patterson, NCSA
Data collection by Maxine Brown.


NORDUnet n*10Gb
Stockholm
Oslo
SURFnet-
IEEAF-
GLORIAD
50Gb
NORDUnet-
GLORIAD 10Gb
Copenhagen
NorthernLight
NORDUnet 50Gb
Amsterdam
NetherLight
Hamburg
UKLight
London
Muenster
CESNET 10Gb
Aachen
Frankfurt
CzechLight
Prague
SURFnet
50Gb
Paris
DAS3-
Grid'5000
10Gb
CERN
Geneva
CERN/TIF
Barcelona
CATLight-I2CAT 10Gb

Network Description Language

- From semantic Web / Resource Description Framework.
- The RDF uses XML as an interchange syntax.
- Data is described by triplets:

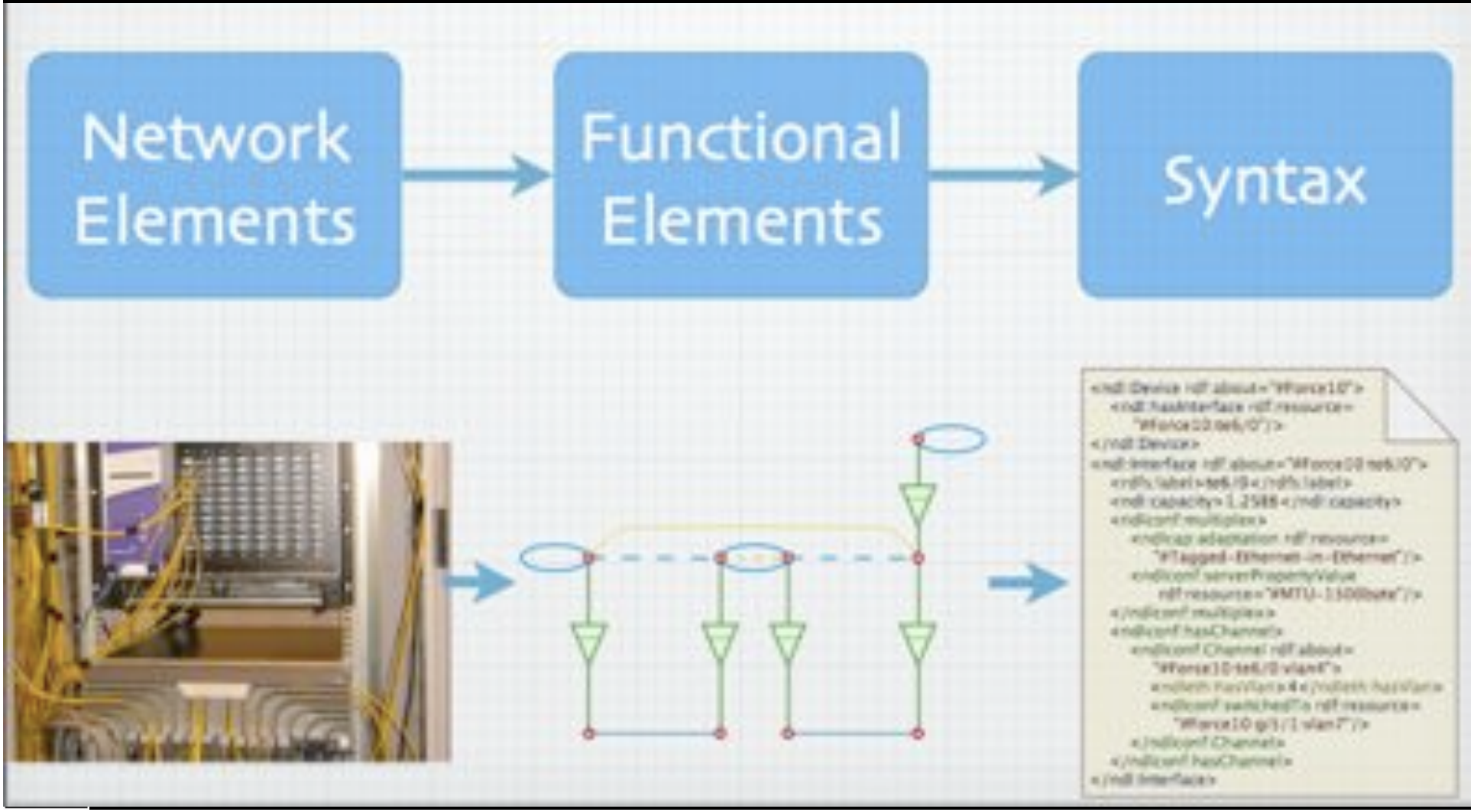


Network Description Language

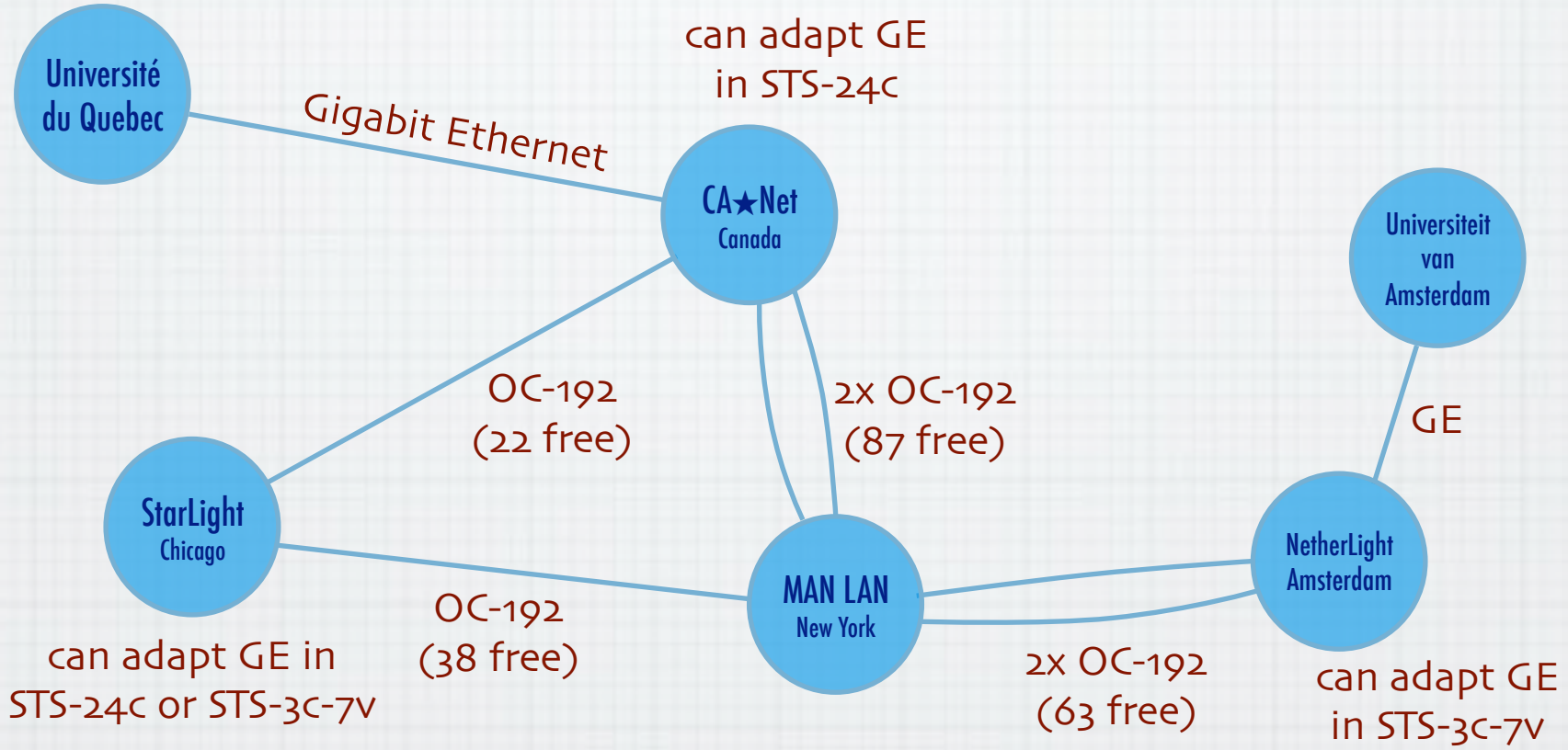
Choice of RDF instead of XML syntax

Grounded modeling based on G0805 description:

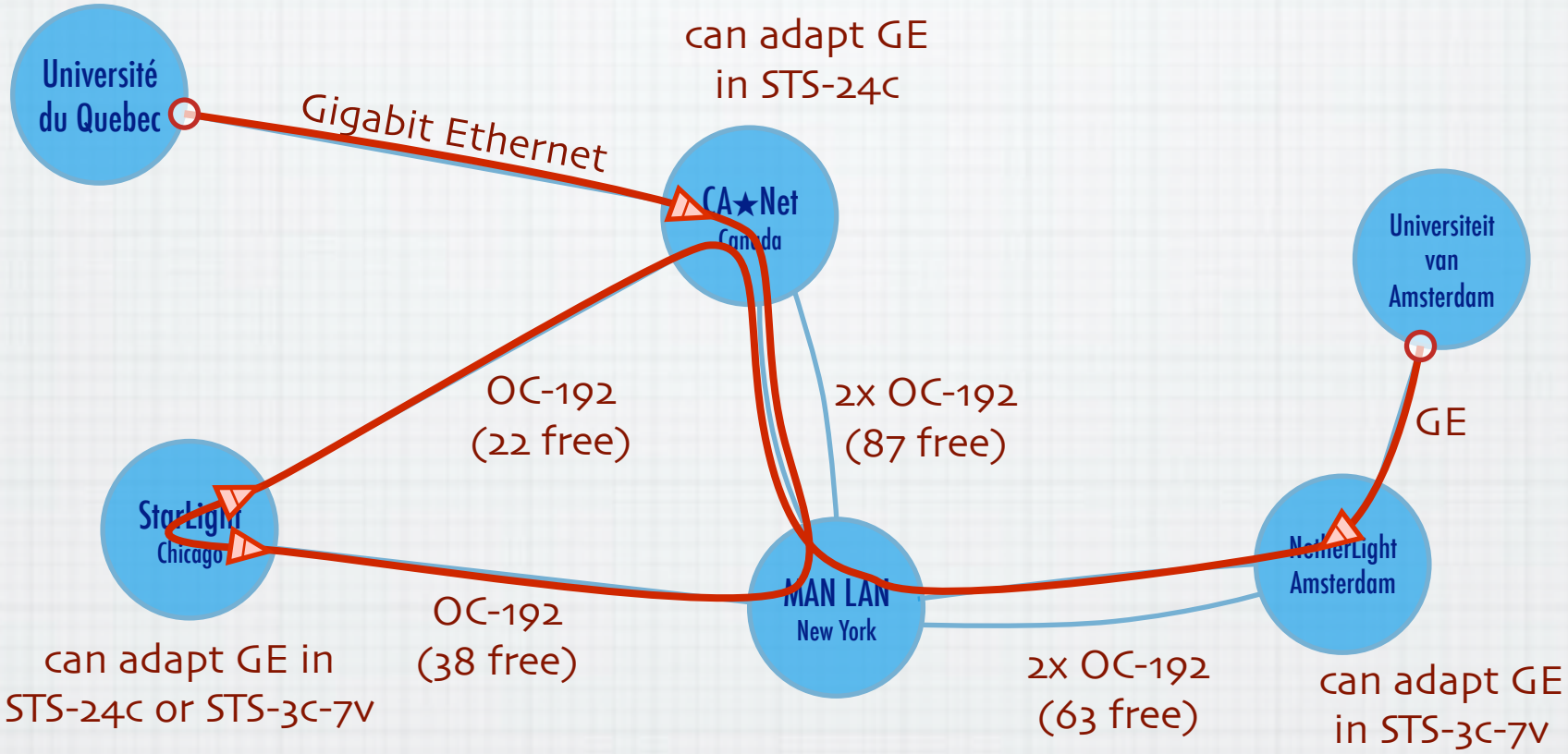
Article: F. Dijkstra, B. Andree, K. Koymans, J. van der Ham, P. Grosso, C. de Laat, *"A Multi-Layer Network Model Based on ITU-T G.805"*



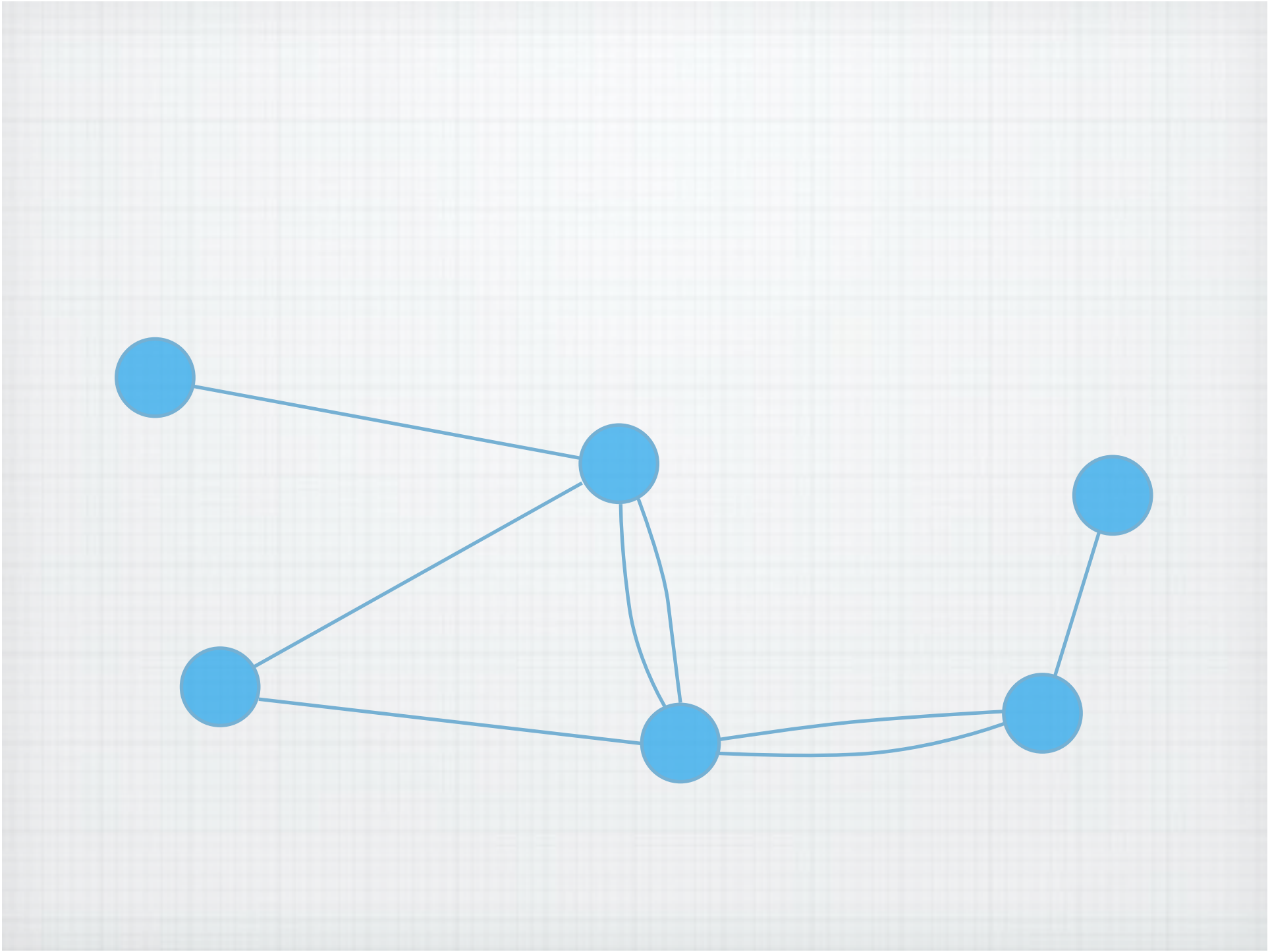
A weird example

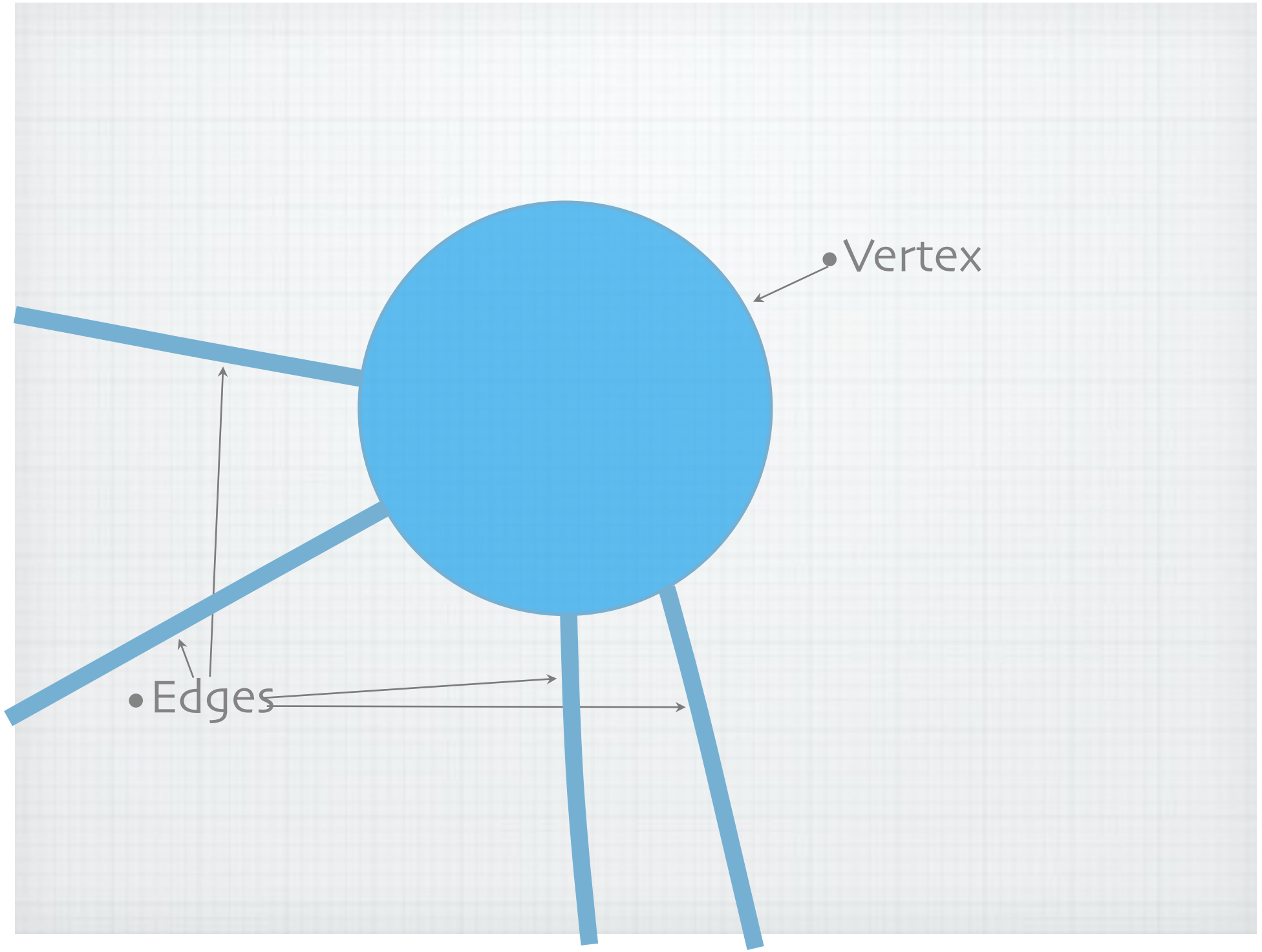


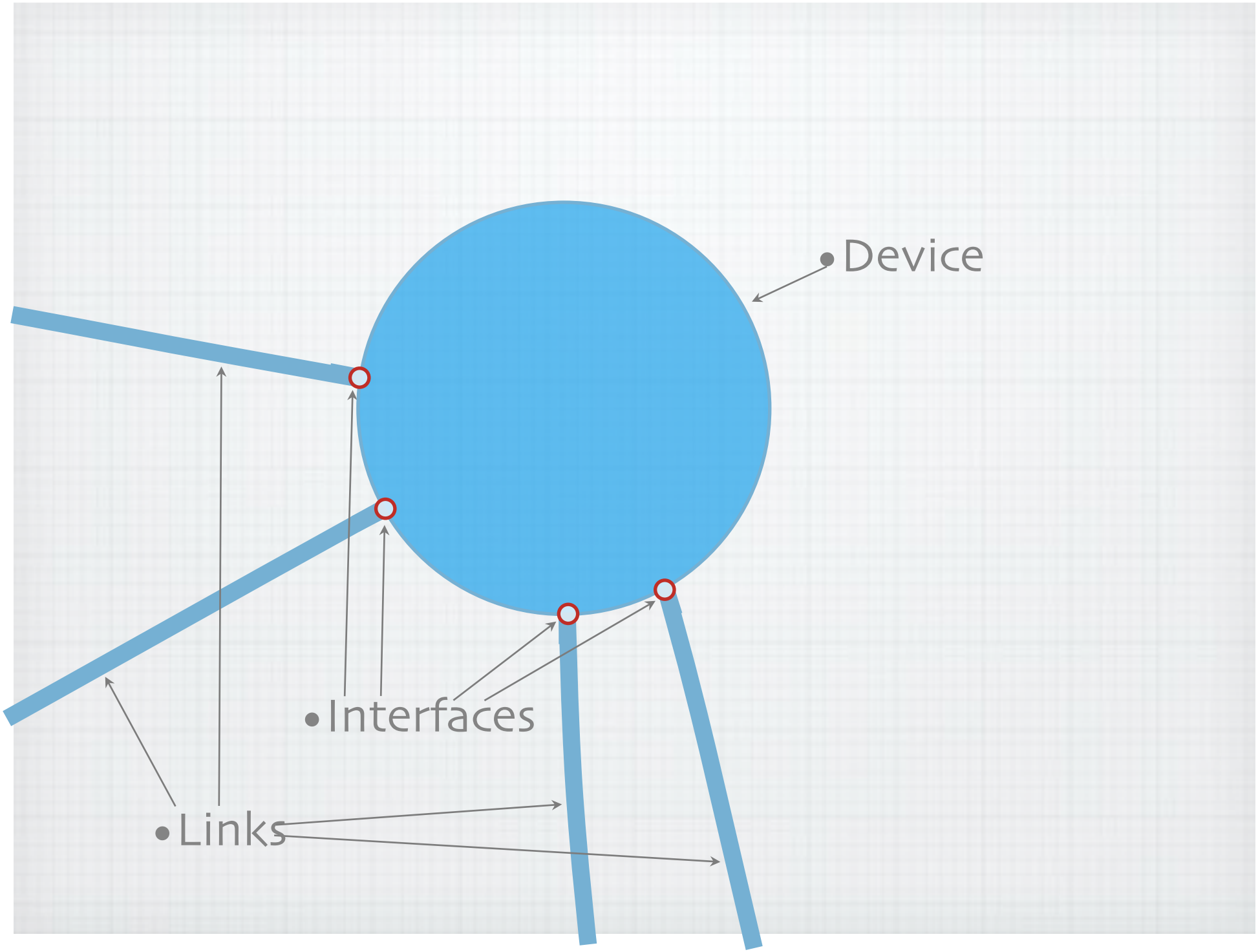
The result :-)

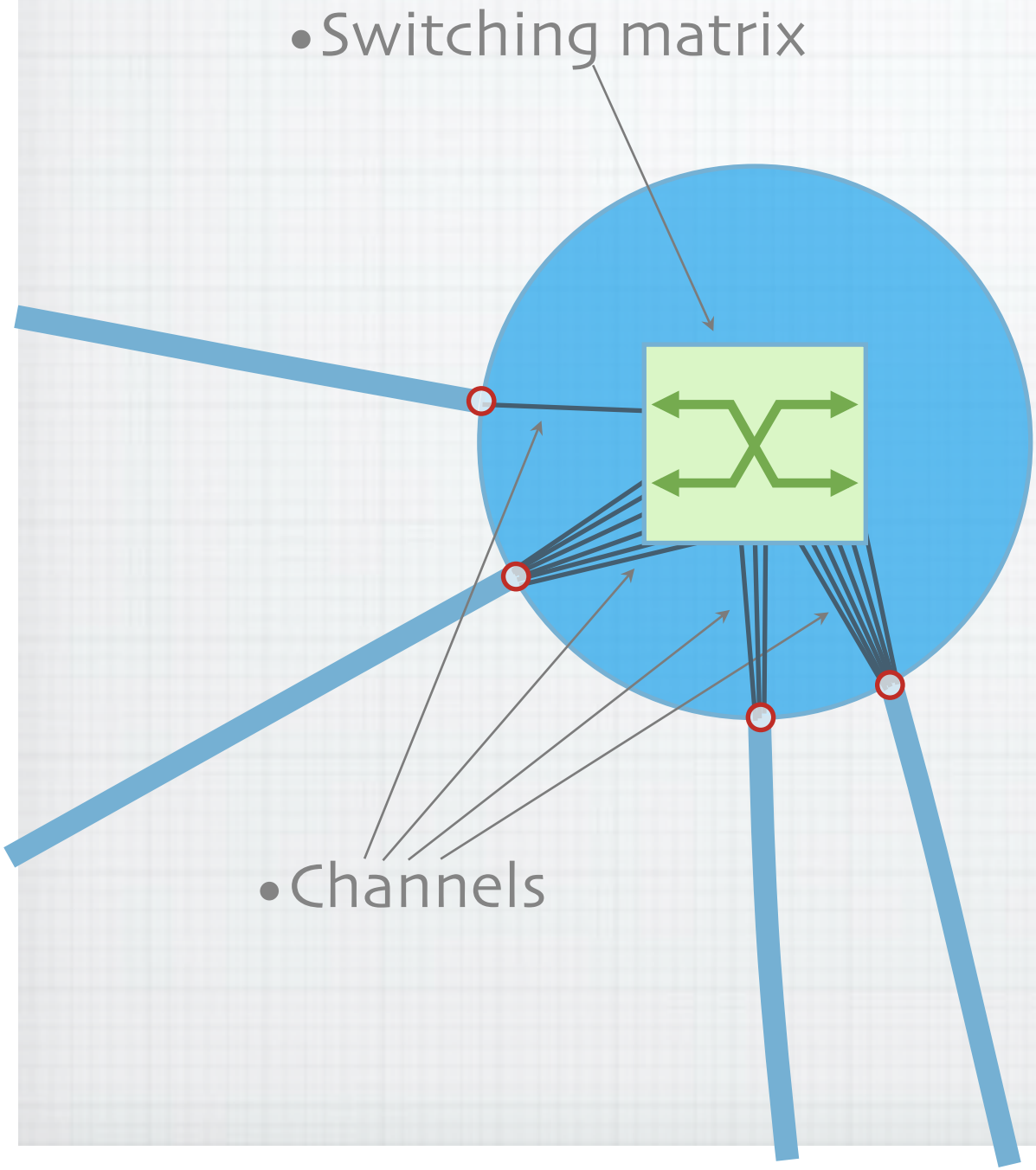


Thanks to Freek Dijkstra & team



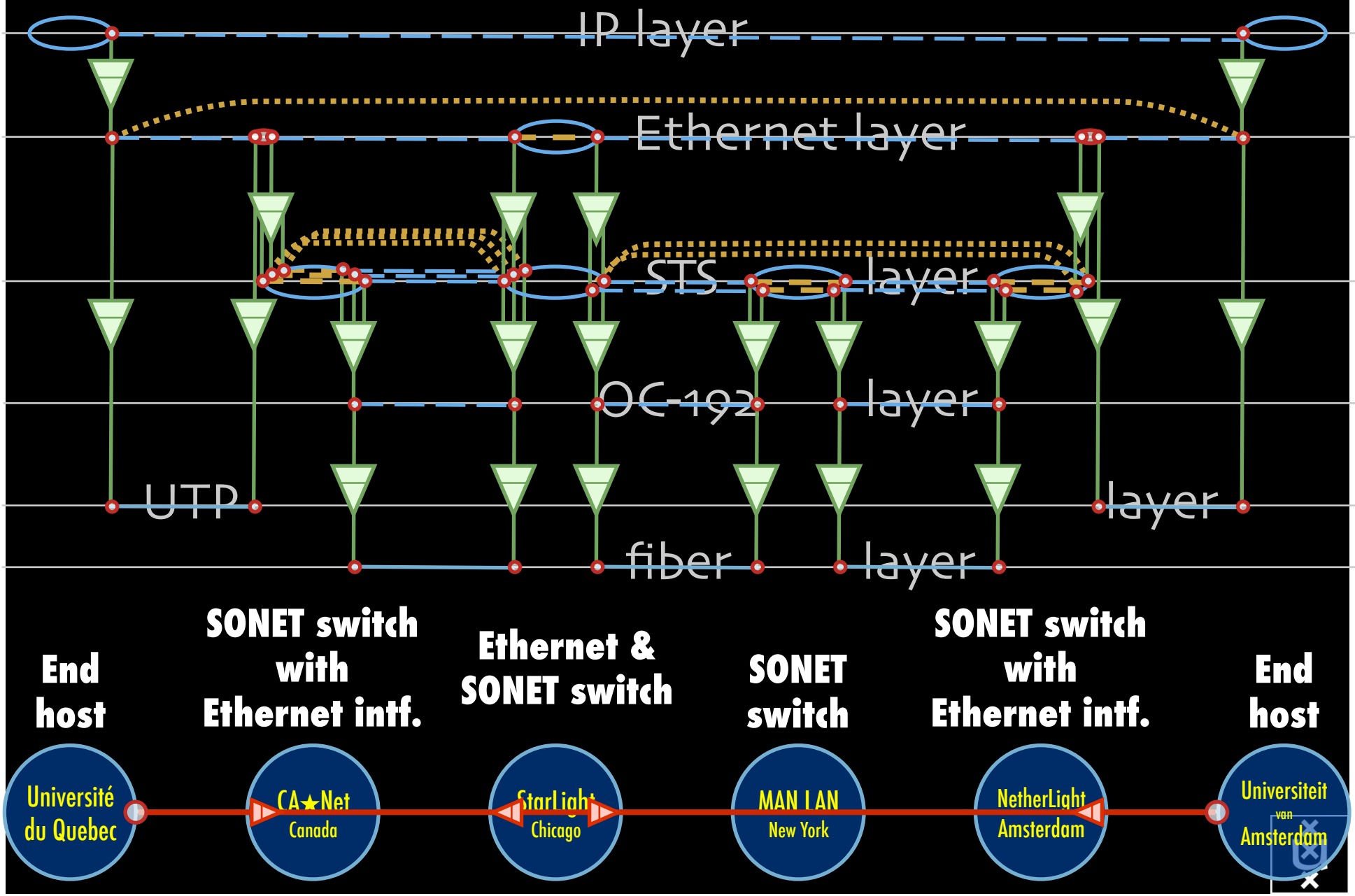

Vertex
Edges


Device
Interfaces
Links



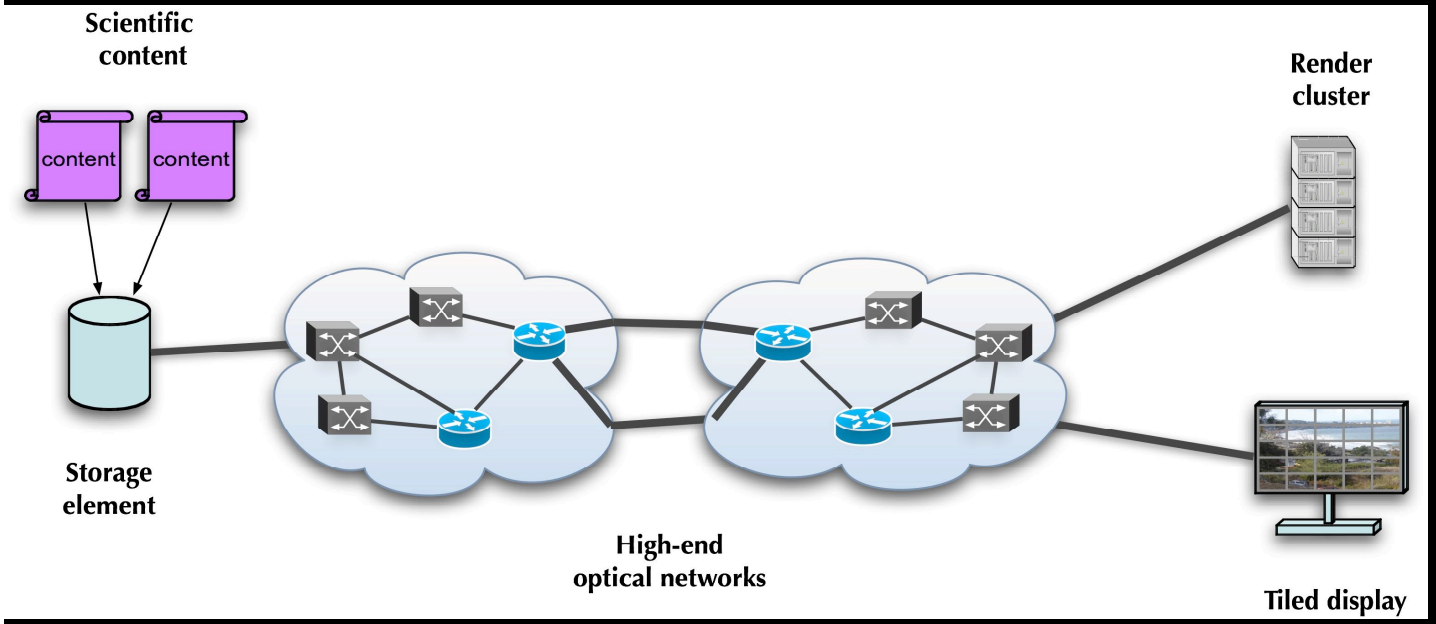
- A devices switches data based on:
  - The source interface
  - One or more labels
- Example label types:
  - Ethernet VLAN
  - SONET STS Channel
  - Wavelength (λ)
- For example, all data from channel 31 of interface 2 is forwarded to channel 28 of interface 4.

Multi-layer extensions to NDL



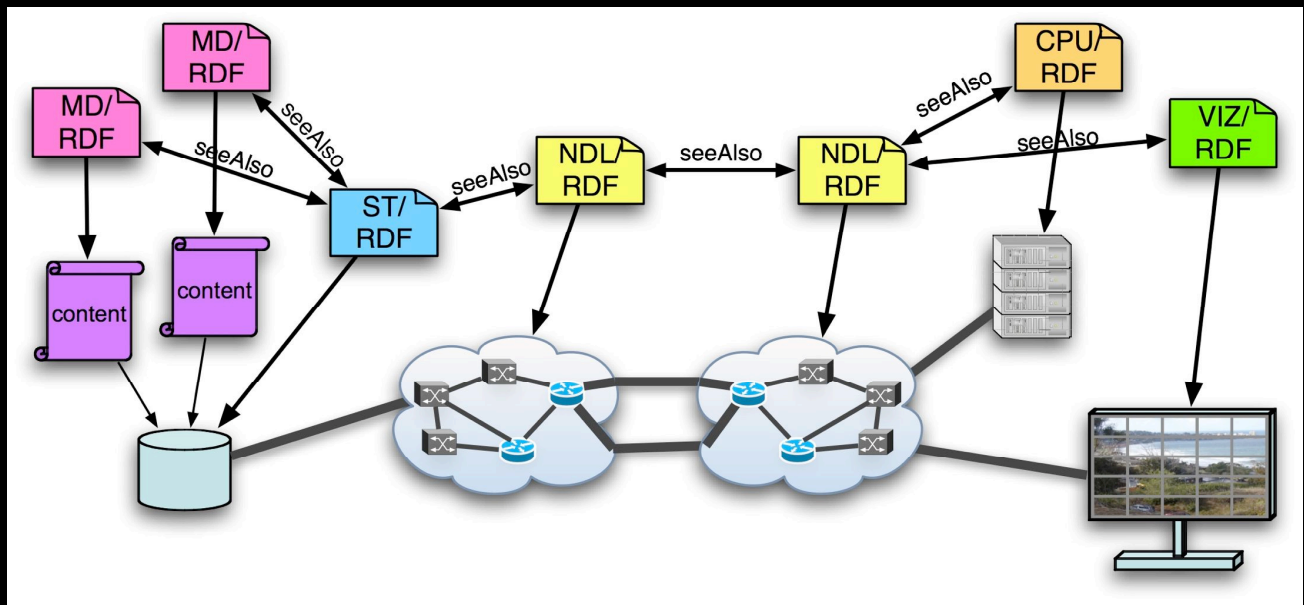

Map
Satellite
Hybrid
Greenland
Canada
United States
North Pacific Ocean
North Atlantic Ocean
Mexico
Cuba
Venezuela
Colombia
Brasil
Peru
Bolivia
Chile
Argentina
Australia
Indonesia
Papua New Guinea
South Pacific Ocean
South Atlantic Ocean
Algeria
Libya
Mauritania
Mali
Niger
Nigeria
Chad
Namibia
Norge
España
Italia
Google

From network to applications

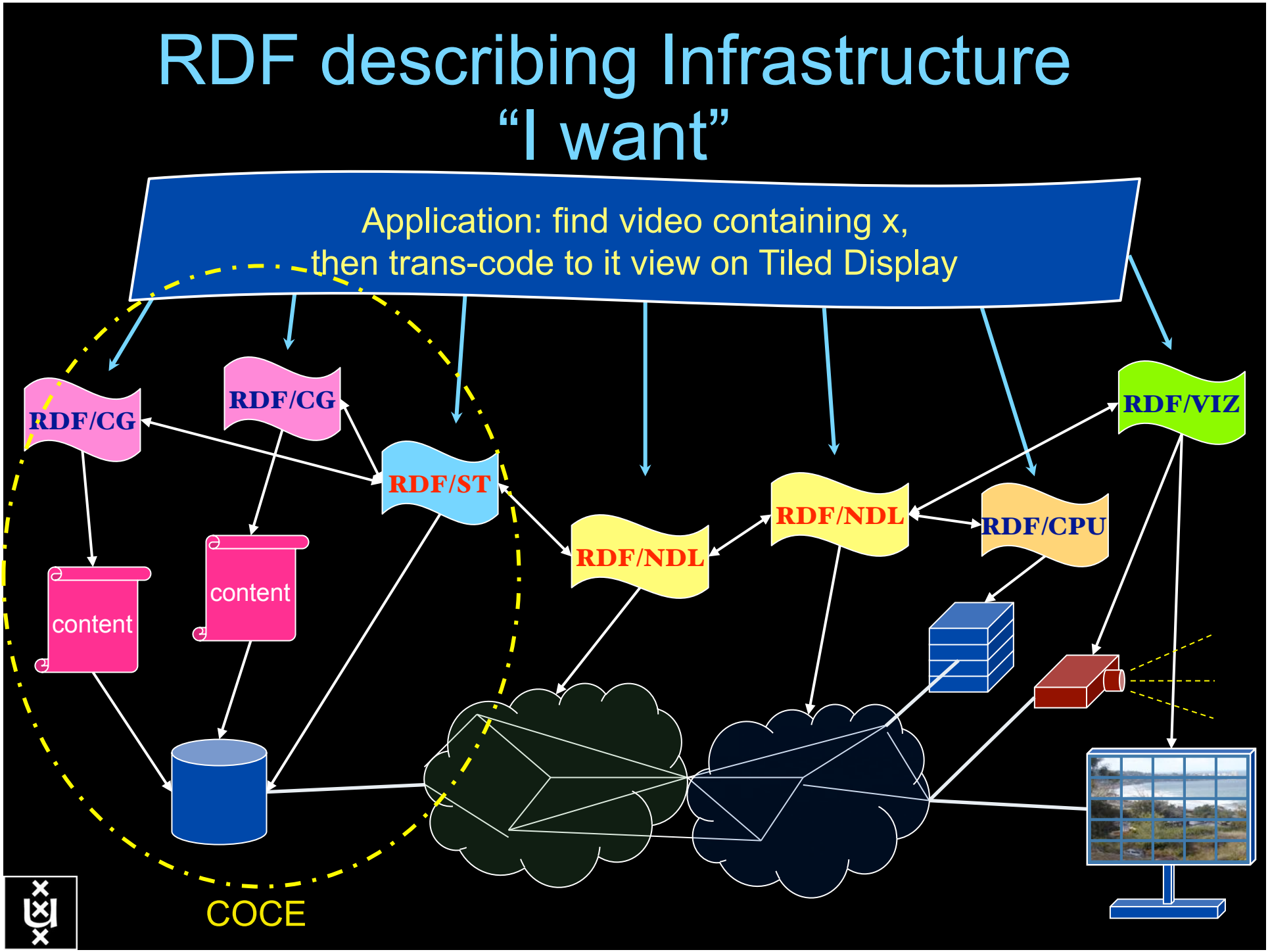


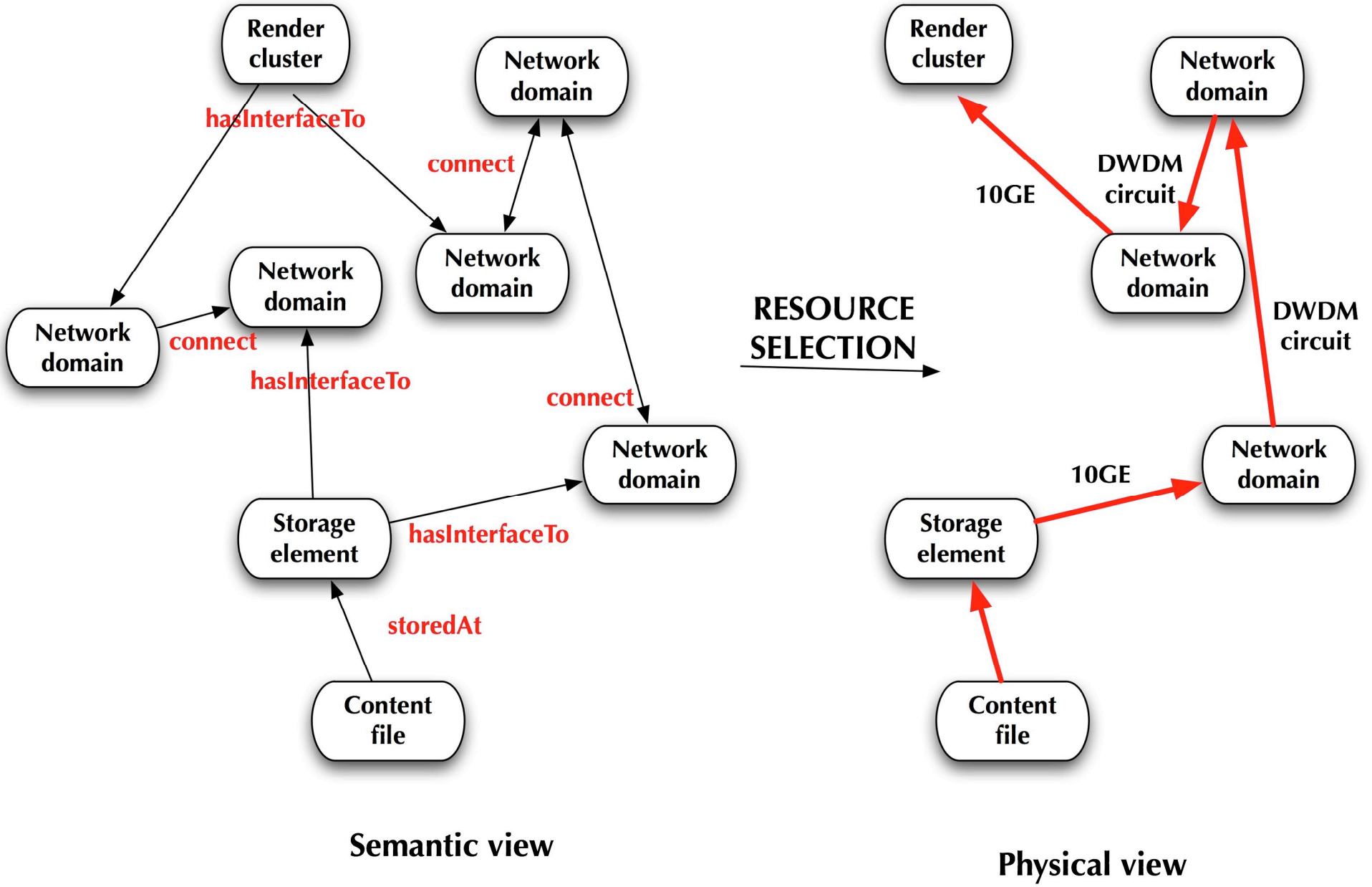
From the physical architecture

To the semantic model for the architecture



RDF describing Infrastructure
"I want"





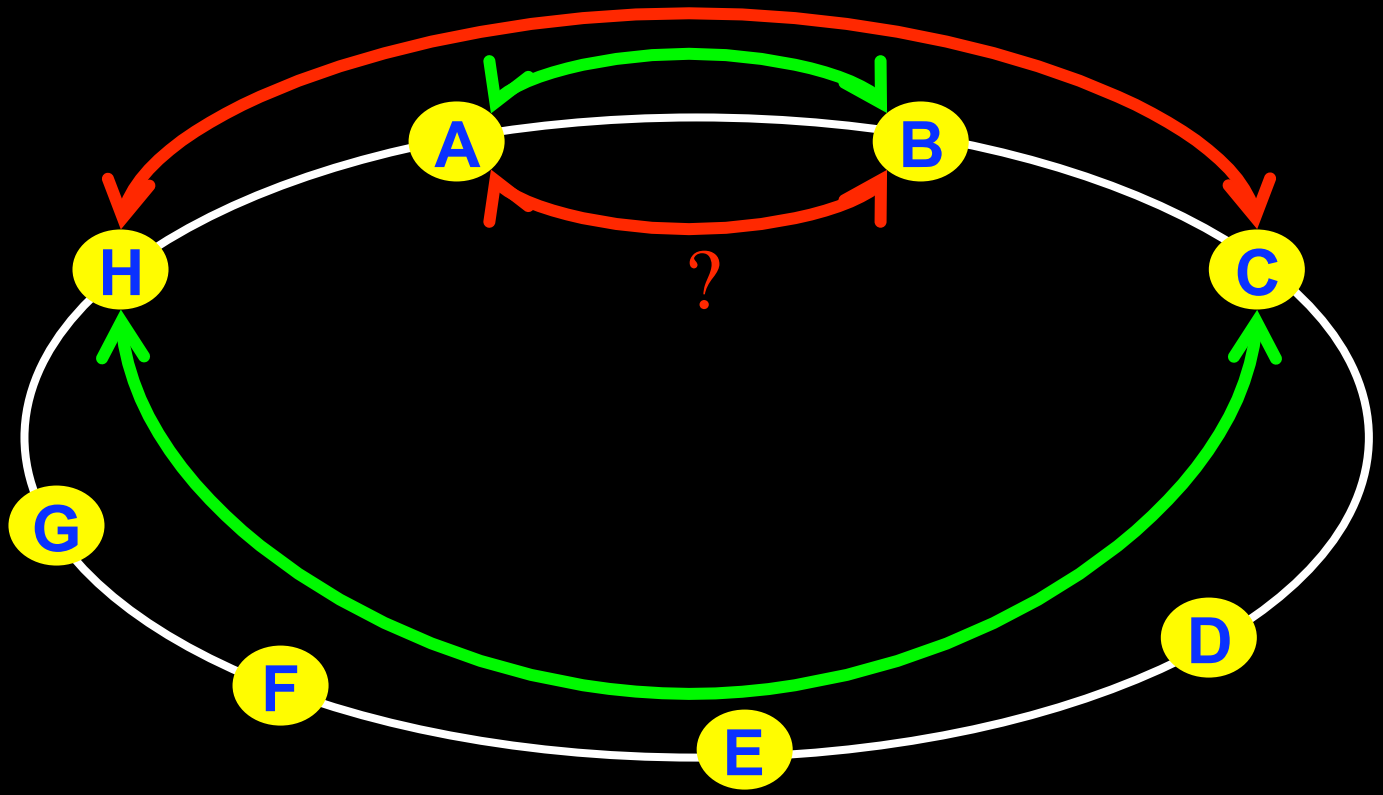
Semantic Reasoning

The Problem

I want HC and AB

Success depends on the order

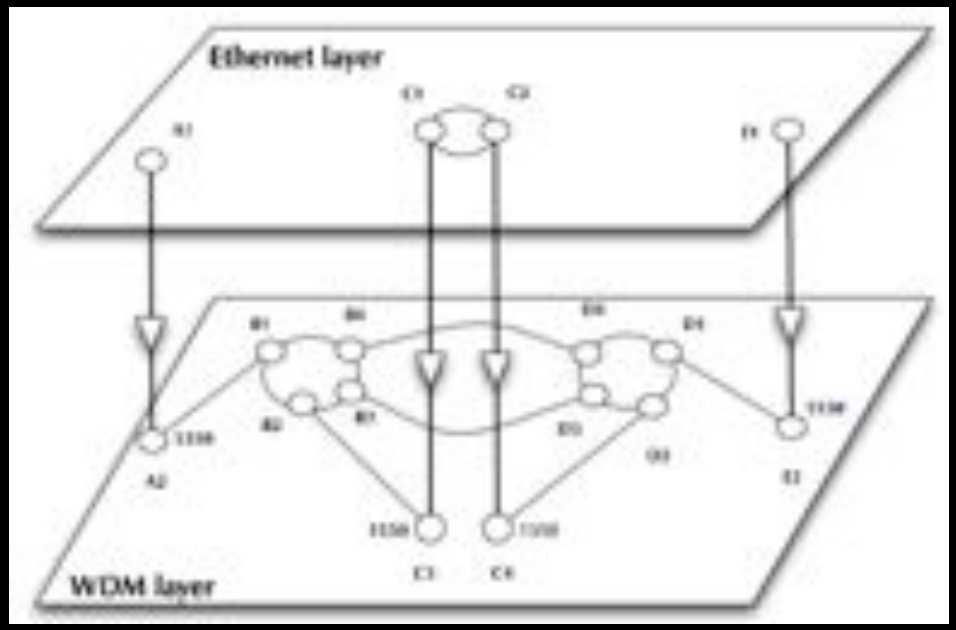
Wouldn't it be nice if I could request [HC, AB, ...]

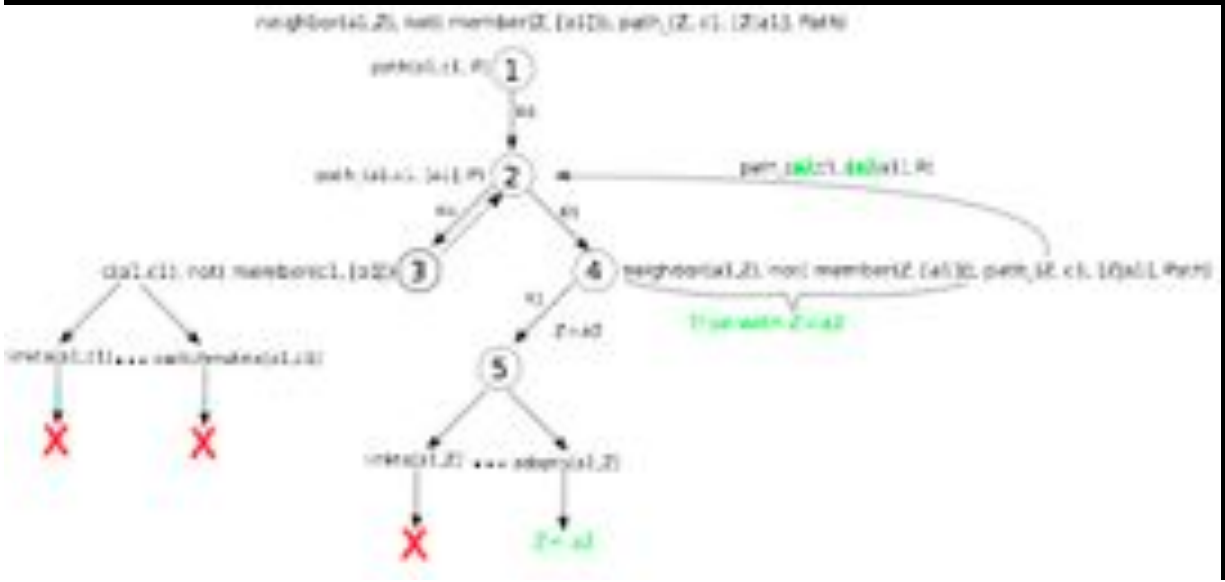


NDL + PROLOG

Research Questions:

- order of requests
- complex requests
- Usable leftovers





- Reason about graphs
- Find sub-graphs that comply with rules

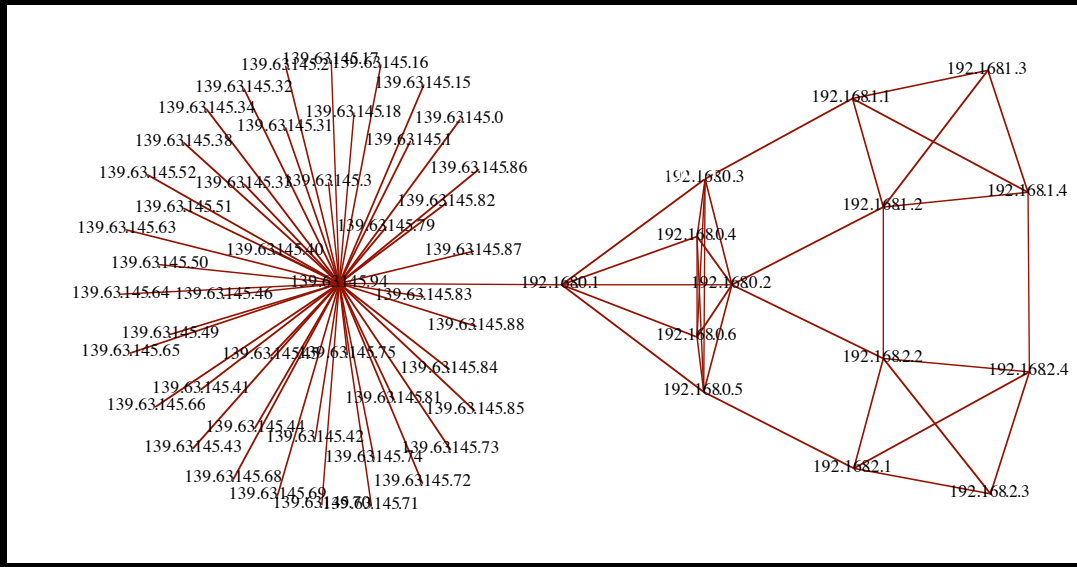
Mathematica enables advanced graph queries, visualizations and real-time network manipulations on UPVNs

Topology matters can be dealt with algorithmically

Results can be persisted using a transaction service built in UPVN

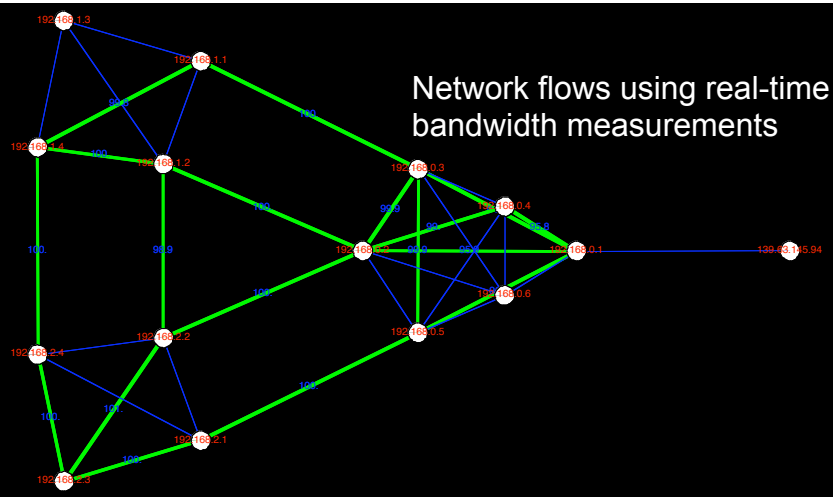
Initialization and BFS discovery of NEs

```
Needs["WebServices`"]
<<DiscreteMath`Combinatorica`
<<DiscreteMath`GraphPlot`
InitNetworkTopologyService["edge.ict.tno.nl"]

Available methods:


{DiscoverNetworkElements,GetLinkBandwidth,GetAllIpLinks,Remote,
NetworkTokenTransaction}

Global`upvnverbose = True;

AbsoluteTiming[nes = BFSDiscover["139.63.145.94"];][[1]]

AbsoluteTiming[result = BFSDiscoverLinks["139.63.145.94", nes];][[1]]

Getting neigbours of: 139.63.145.94
Internal links: {192.168.0.1, 139.63.145.94}
(...)
Getting neigbours of:192.168.2.3
```



Transaction on shortest path with tokens

```
nodePath = ConvertIndicesToNodes[
           ShortestPath[     g,
                        Node2Index[nids,"192.168.3.4"],
                        Node2Index[nids,"139.63.77.49"]],
                        nids];
Print["Path: ", nodePath];
If[NetworkTokenTransaction[nodePath, "green"]==True,
      Print["Committed"], Print["Transaction failed"]];

Path:
{192.168.3.4,192.168.3.1,139.63.77.30,139.63.77.49}

Committed
```



Network flows using real-time bandwidth measurements

ref: Robert J. Meijer, Rudolf J. Strijkers, Leon Gommans, Cees de Laat, User Programmable Virtualiized Networks, accepted for publication to the IEEE e-Science 2006 conference Amsterdam.

StarPlane

TouchTable Demonstration @ SC08

Interactive programmable networks

OGF NML-WG

Open Grid Forum - Network Markup Language workgroup

Chairs:

- Paola Grosso – Universiteit van Amsterdam
- Martin Swany – University of Delaware

Purpose:

To describe network topologies, so that the outcome is a standardized network description ontology and schema, facilitating interoperability between different projects.

https://forge.gridforum.org/sf/projects/nml-wg

Questions ?

Accepted paper: A Declarative Approach to Multi-Layer Path Finding Based on Semantic Network Descriptions.

Not on the memory stick, so:

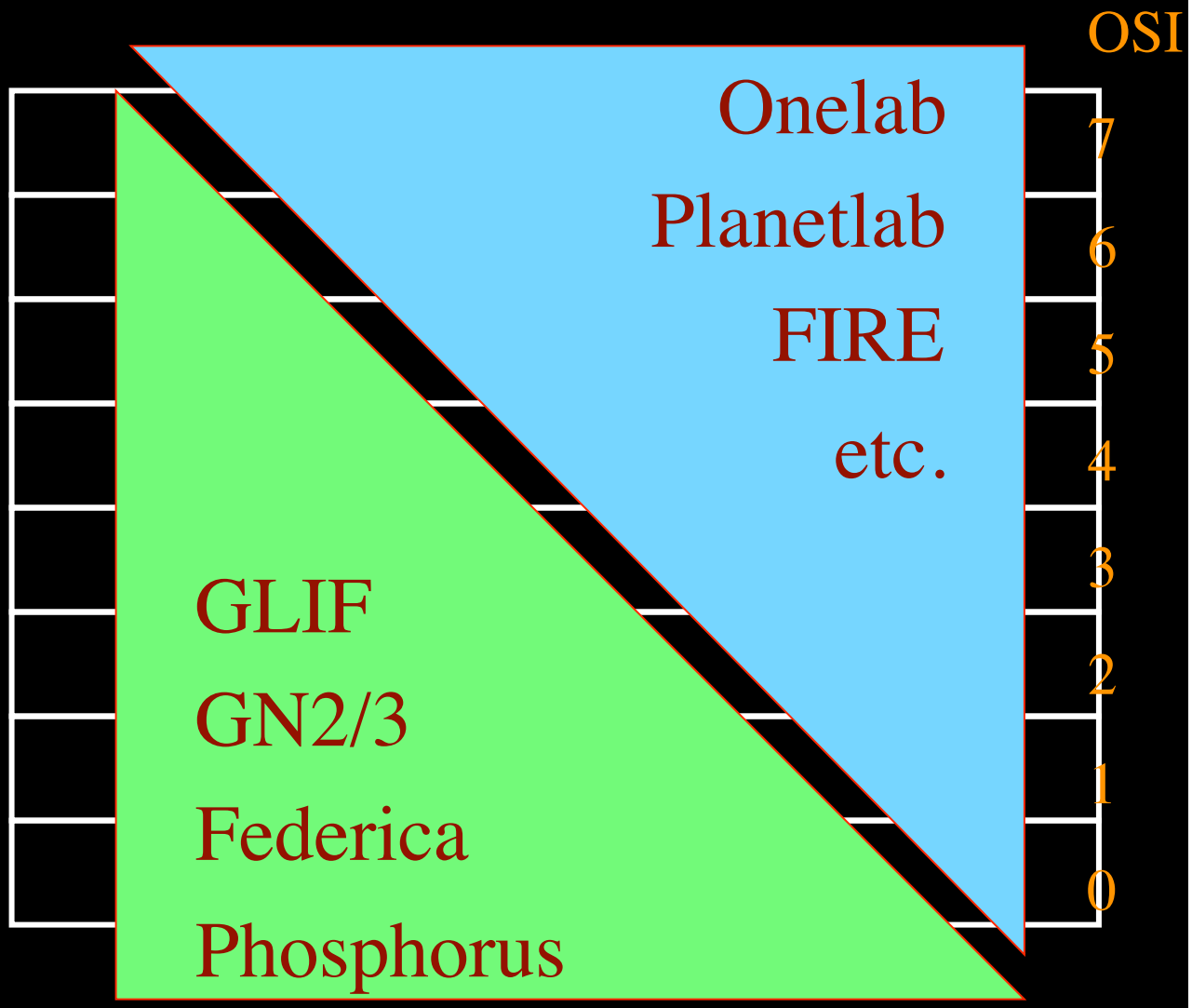
http://delaat.net:/~delaat/papers/declarative_path_finding.pdf

Thanks: Paola Grosso & Jeroen vd Ham & Freek Dijkstra & team for several of the slides.

SURF NET

My view

- needs repeatable experiment
- needs QoS & lightpaths
- needs capacity and capability
- needs infrastructure descriptions



TeraThinking

- What constitutes a Tb/s network?
- CALIT2 has 8000 Gigabit drops ?->? Terabit Lan?
- look at 80 core Intel processor
  - cut it in two, left and right communicate 8 TB/s
- think back to teraflop computing!
  - MPI makes it a teraflop machine
- massive parallel channels in hosts, NIC's
- TeraApps programming model supported by
  - TFlops -> MPI / Globus
  - TBytes -> OGSA/DAIS
  - TPixels -> SAGE
  - TSensors -> LOFAR, LHC, LOOKING, CineGrid, ...
  - Tbit/s -> ?

ref Larry Smarr & CdL

Multi Layer Service Architecture