Challenges in Enabling Grid Computing over Optical Networks

Cees de Laat

GLIF.is founding member

SURFnet

EU

BSIK

NWO

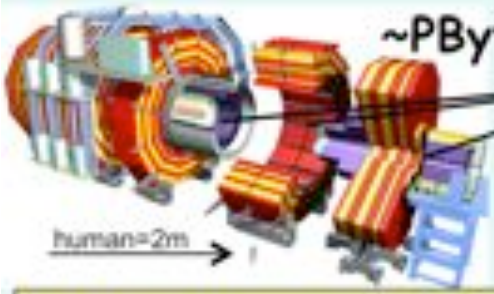
University of Amsterdam

TNO

NCF

LHC Data Grid Hierarchy

CMS as example, Atlas is similar



100000 flops/byte

~PByte/sec

Online System

~100 MBytes/sec

Tier 0 +1

HPSS

event reconstruction

event simulation

CMS detector: 15m X 15m X 22m

12,500 tons, $700M.

~2.5 Gbits/sec

Tier 1

Italian Regional Center

German Regional Center

NIKHEF Dutch Regional Center

FermiLab, USA Regional Center

HPSS

...

~0.6-2.5 Gbps

analysis

Tier2 Center

Tier 2

~0.6-2.5 Gbps

Tier 3

Institute ~0.25TIPS

Physics data cache

100 - 1000 Mbits/sec

Tier 4

Workstations

Courtesy Harvey Newman, CalTech and CERN

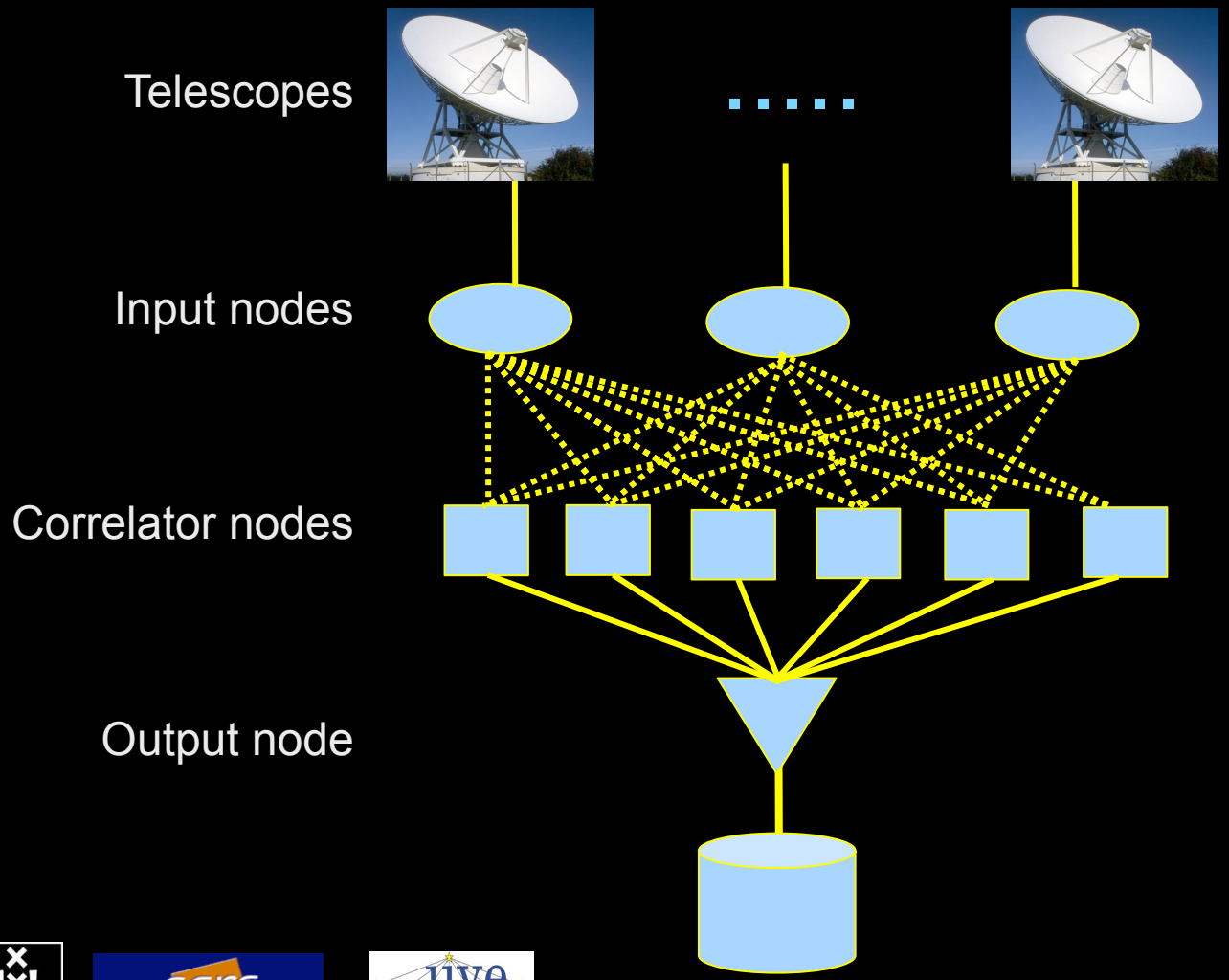
CERN/CMS data goes to 6-8 Tier 1 regional centers, and from each of these to 6-10 Tier 2 centers.

Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

2000 physicists in 31 countries are involved in this 20-year experiment in which DOE is a major player.

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

The “Dead Cat” demo

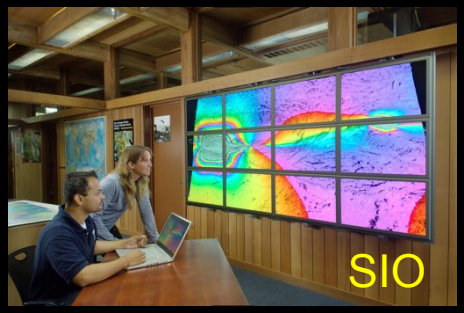
SC2004 & iGrid2005

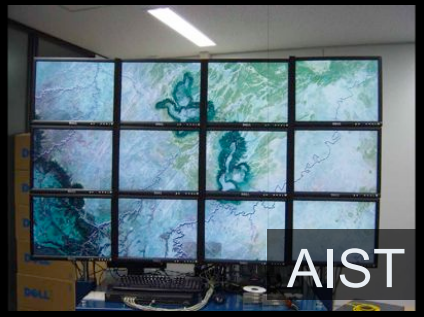
SC2004,
Pittsburgh,
Nov. 6 to 12, 2004
iGrid2005,
San Diego,
sept. 2005

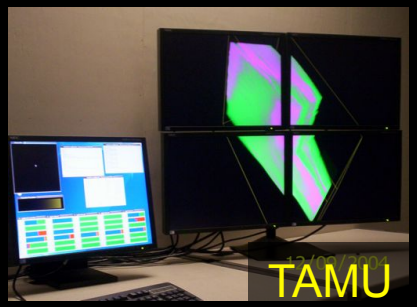
Produced by:
Michael Scarpa
Robert Belleman
Peter Sloot

Many thanks to:
AMC
SARA
GigaPort
UvA/AIR
Silicon Graphics, Inc.
Zoölogisch Museum

US and International OptIPortal Sites





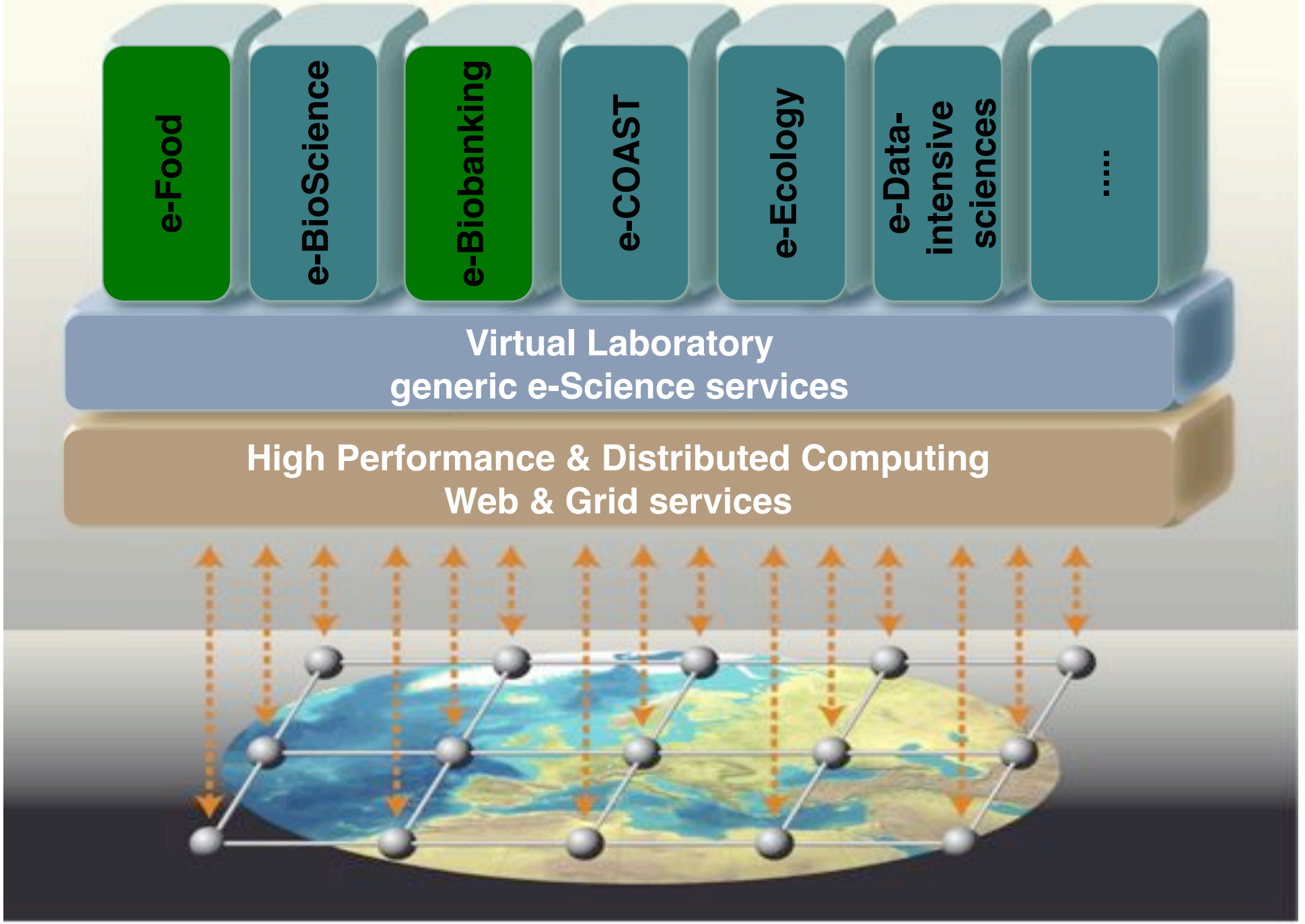


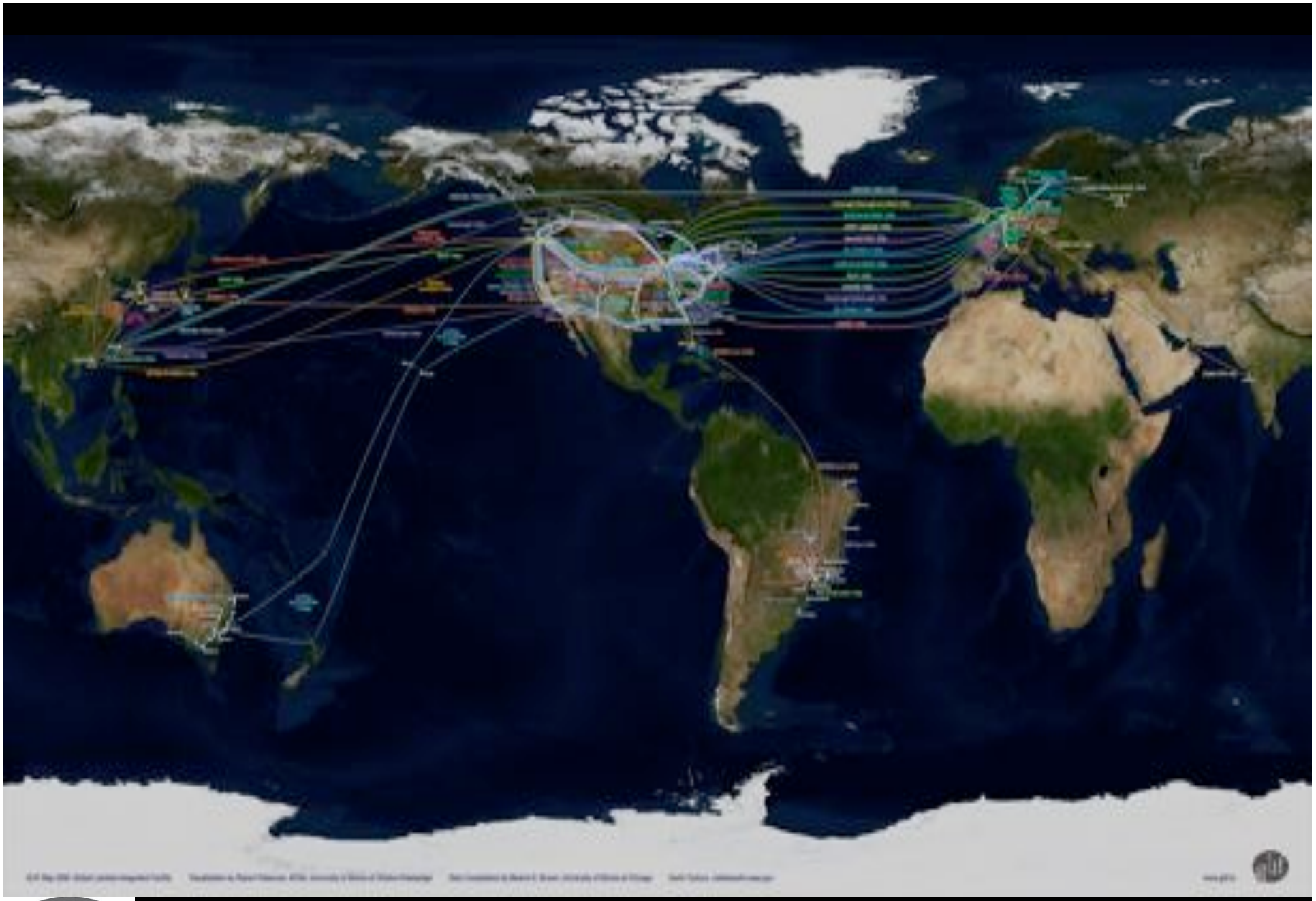
IJKDIJK

Sensor grid: instrument the dikes

First controlled breach occurred on sept 27th ‘08:

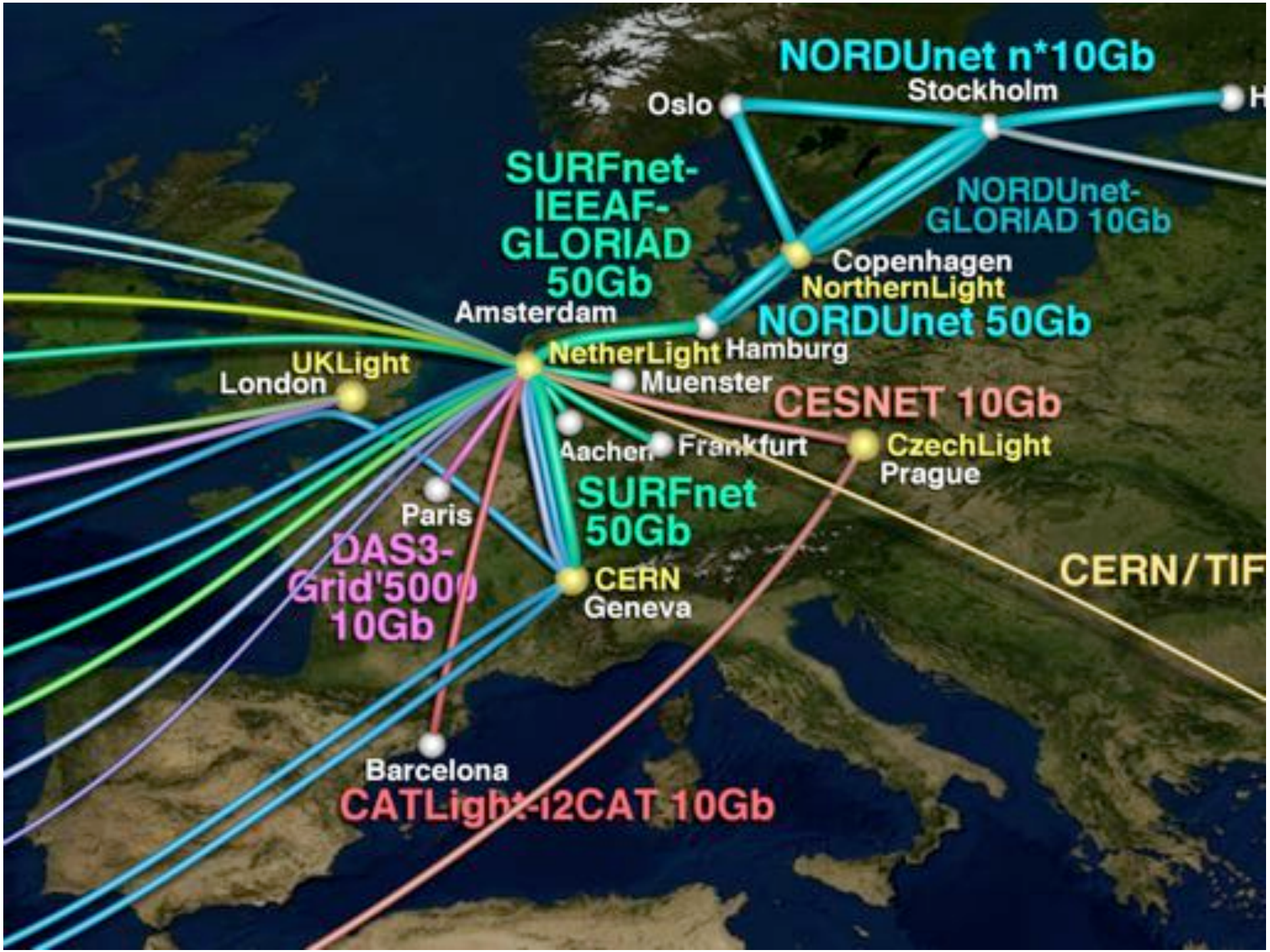
30000 sensors (microphones) to cover all Dutch dikes

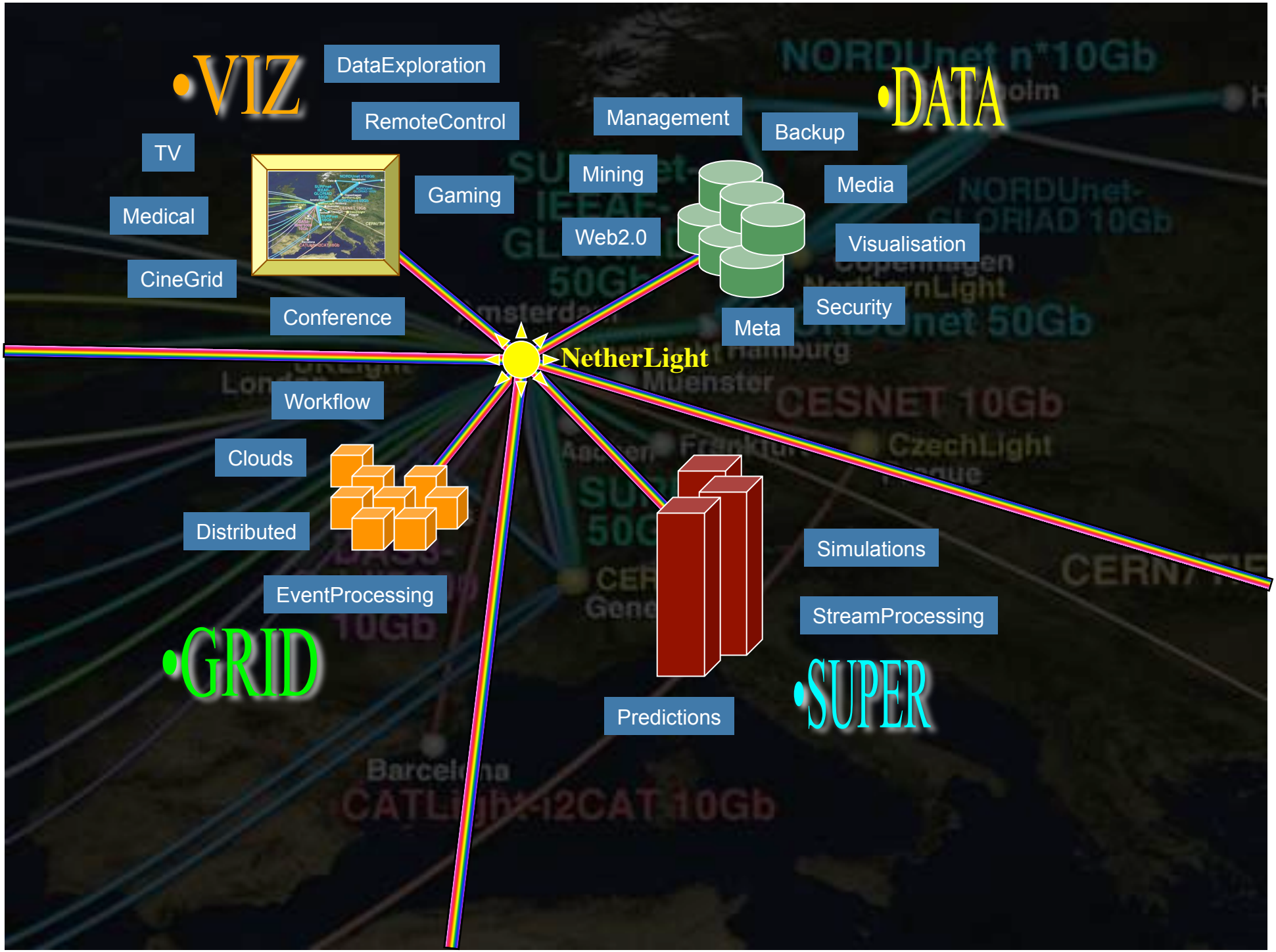

e-Food
e-BioScience
e-Biobanking
e-COAST
e-Ecology
e-Data-intensive sciences
.....
Virtual Laboratory
generic e-Science services
High Performance & Distributed Computing
Web & Grid services



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


•VIZ
DataExploration
RemoteControl
TV
Medical
CineGrid
Conference
Gaming
•DATA
Management
Backup
Mining
Media
Web2.0
Visualisation
Security
Meta
NetherLight
Workflow
Clouds
Distributed
EventProcessing
•GRID
Simulations
StreamProcessing
Predictions
•SUPER

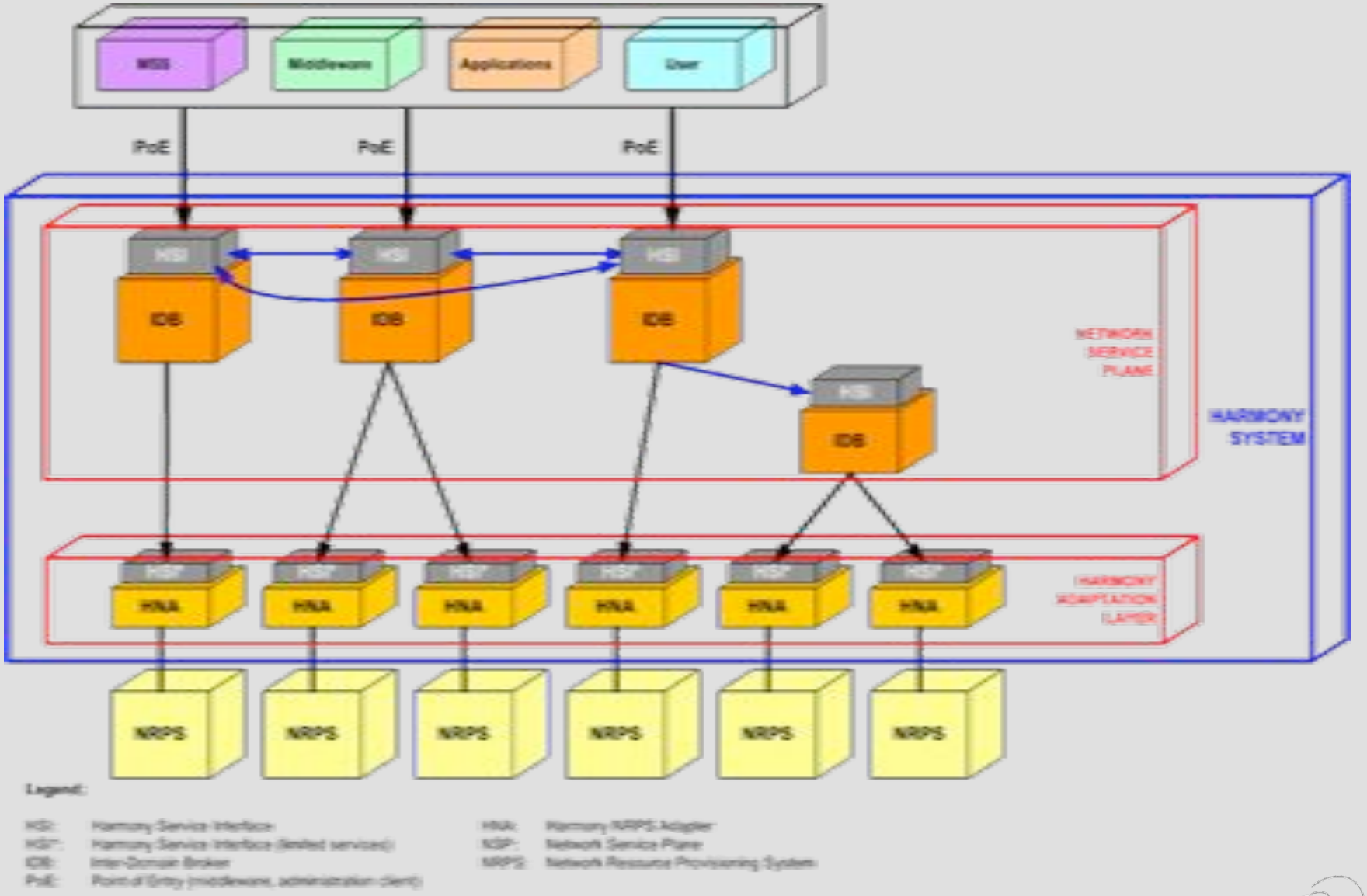
Who (just to name a few) working Interfaces - similar solutions addressing the previous challenges

- CANARIE - UCLP
- ESnet = OSCARS
- G-LAMBDA - GNS-WSI
- HPDM - VLAN based lightpaths
- NORTEL - DRAC
- UvA - Token based service, NDL, etc
- Phosphorous - G2MPLS, UCLP2, NRPS - > NSP
- GN2 - JRA3 - AutoBahn - IDM
- I2 - DRAGON/ HOPI - DCN
- FermiLab - Lambda Station
- DoE - LambdaStation, TeraPaths

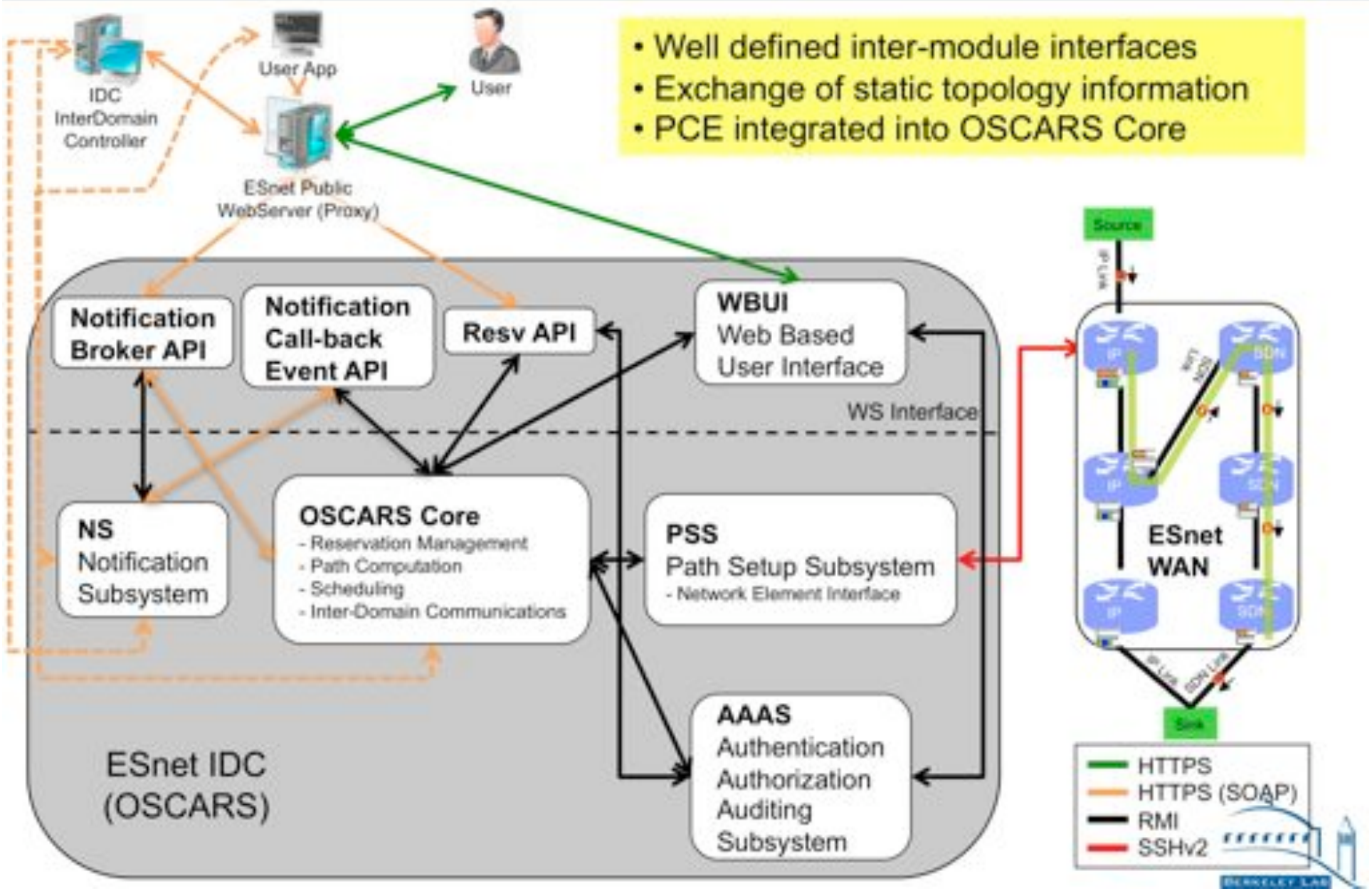
MCNC

Slide: Gigi Karmous Edwards

Harmony architecture (I)



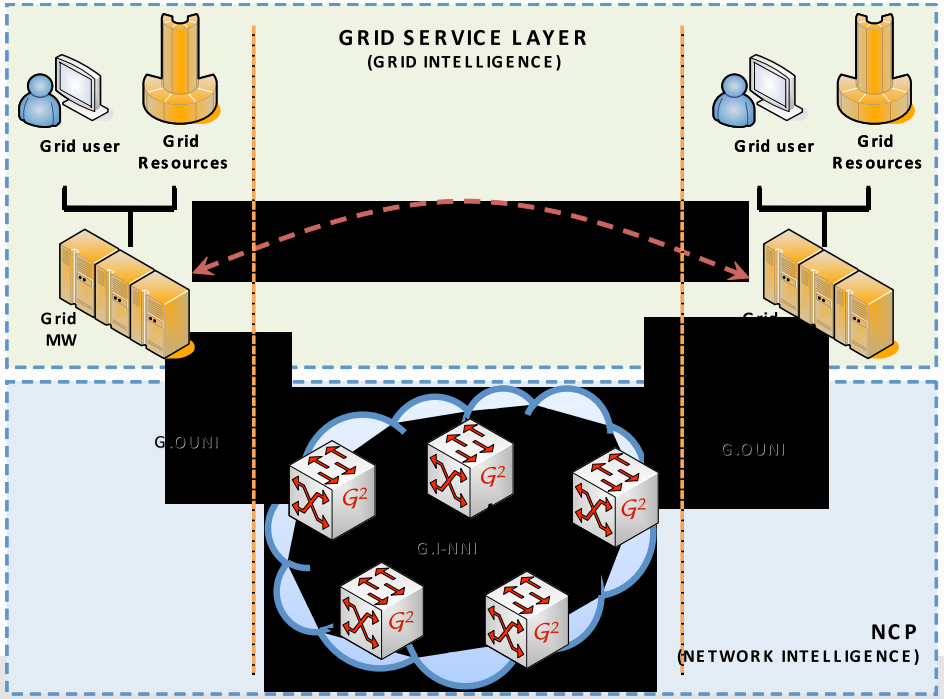
OSCARS Current (v0.5) Implementation



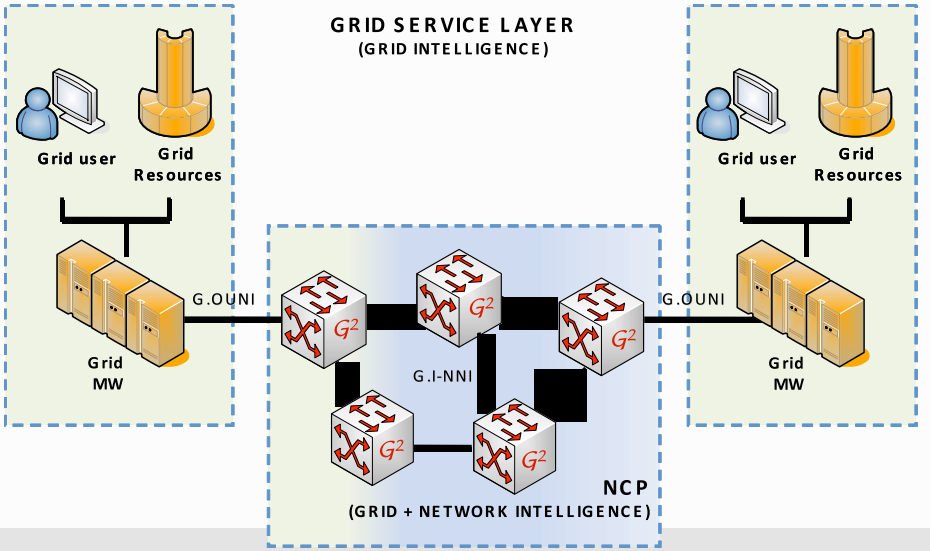
Grid-aware GMPLS (G²MPLS) for Grid Network Services

- Two models for the layering between Grid and Network resources

G²MPLS overlay model



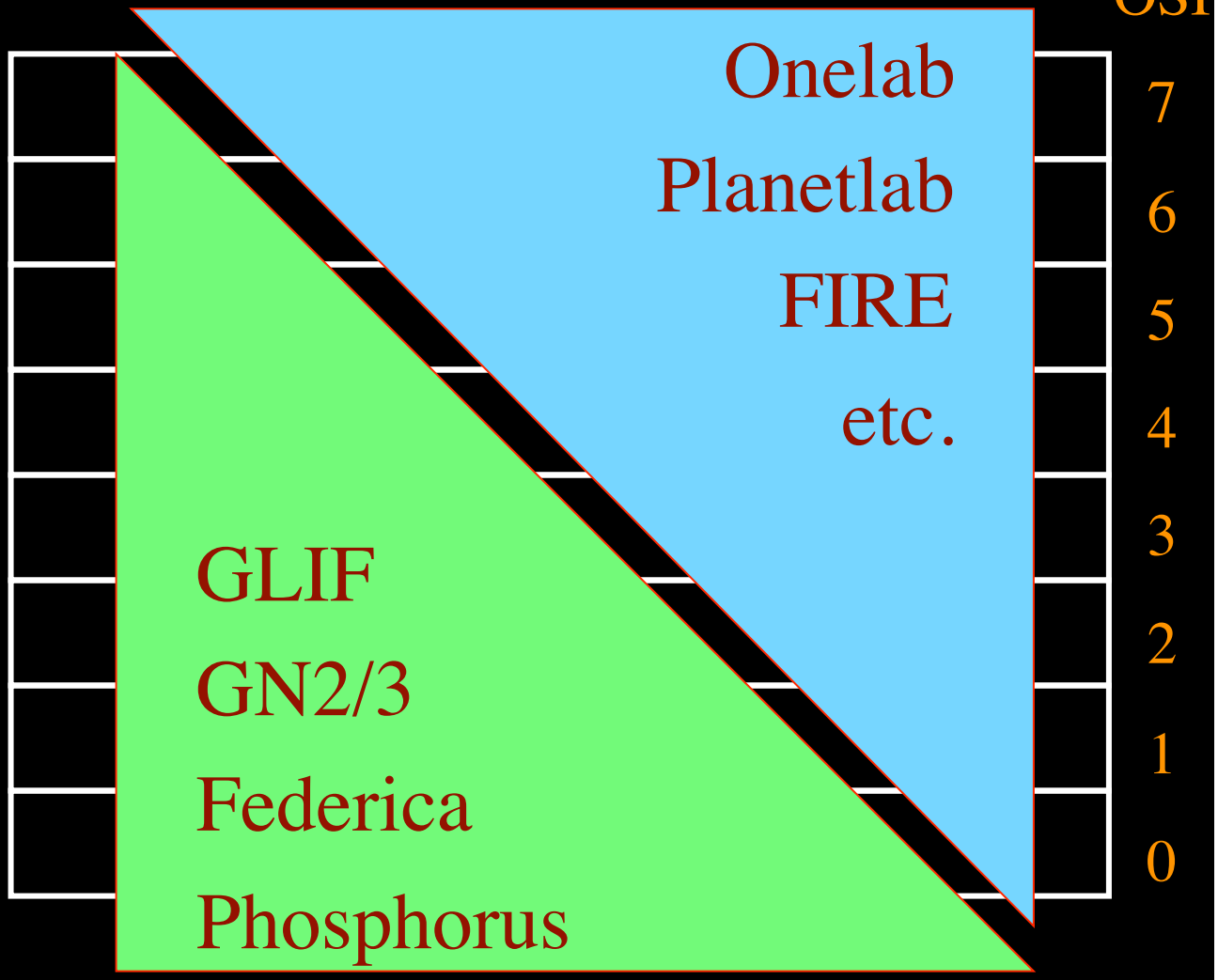
G²MPLS integrated model



different scope with respect to the IETF GMPLS Overlay & Peer

My view

- needs repeatable experiments
- 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 turns a room full of pc's in 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 -> ?

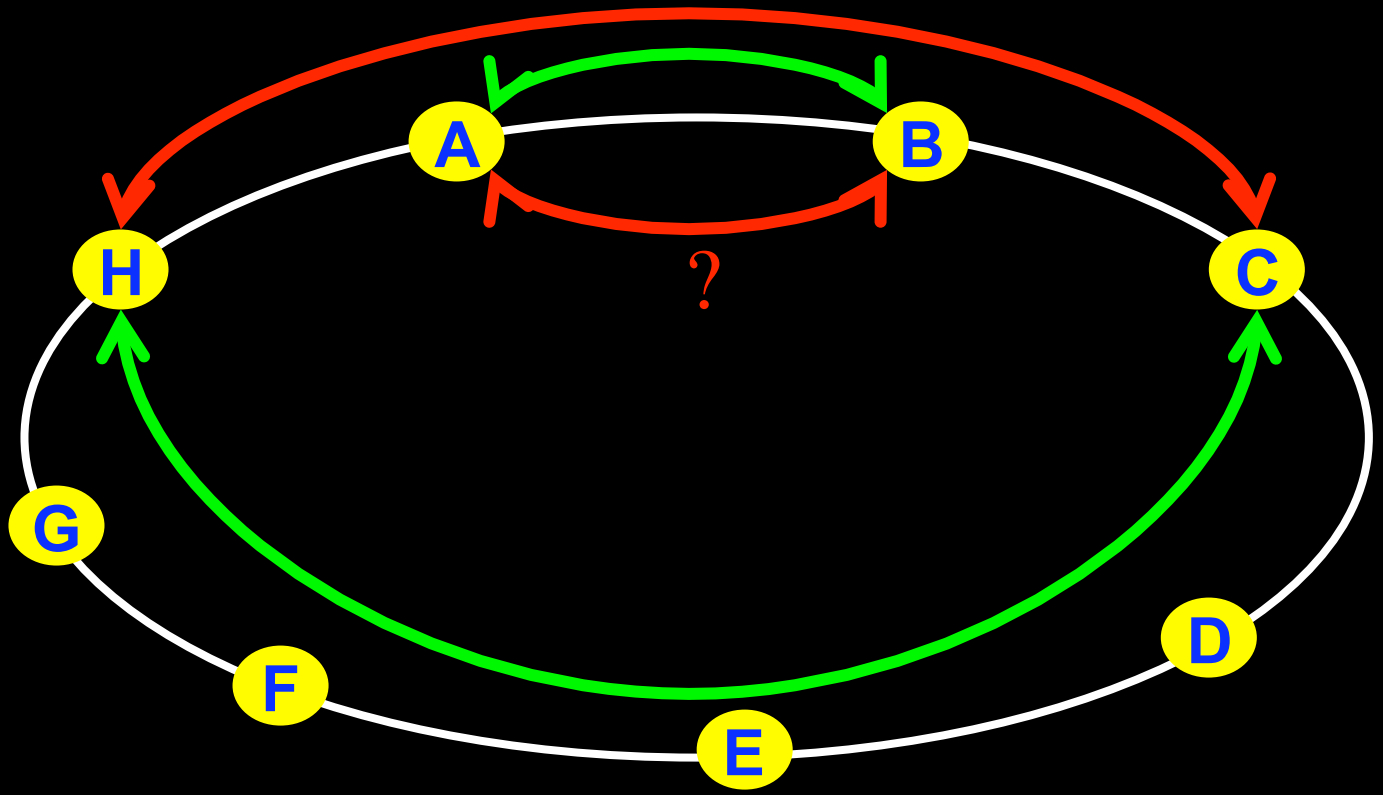
ref Larry Smarr & CdL

The Problem

I want HC and AB

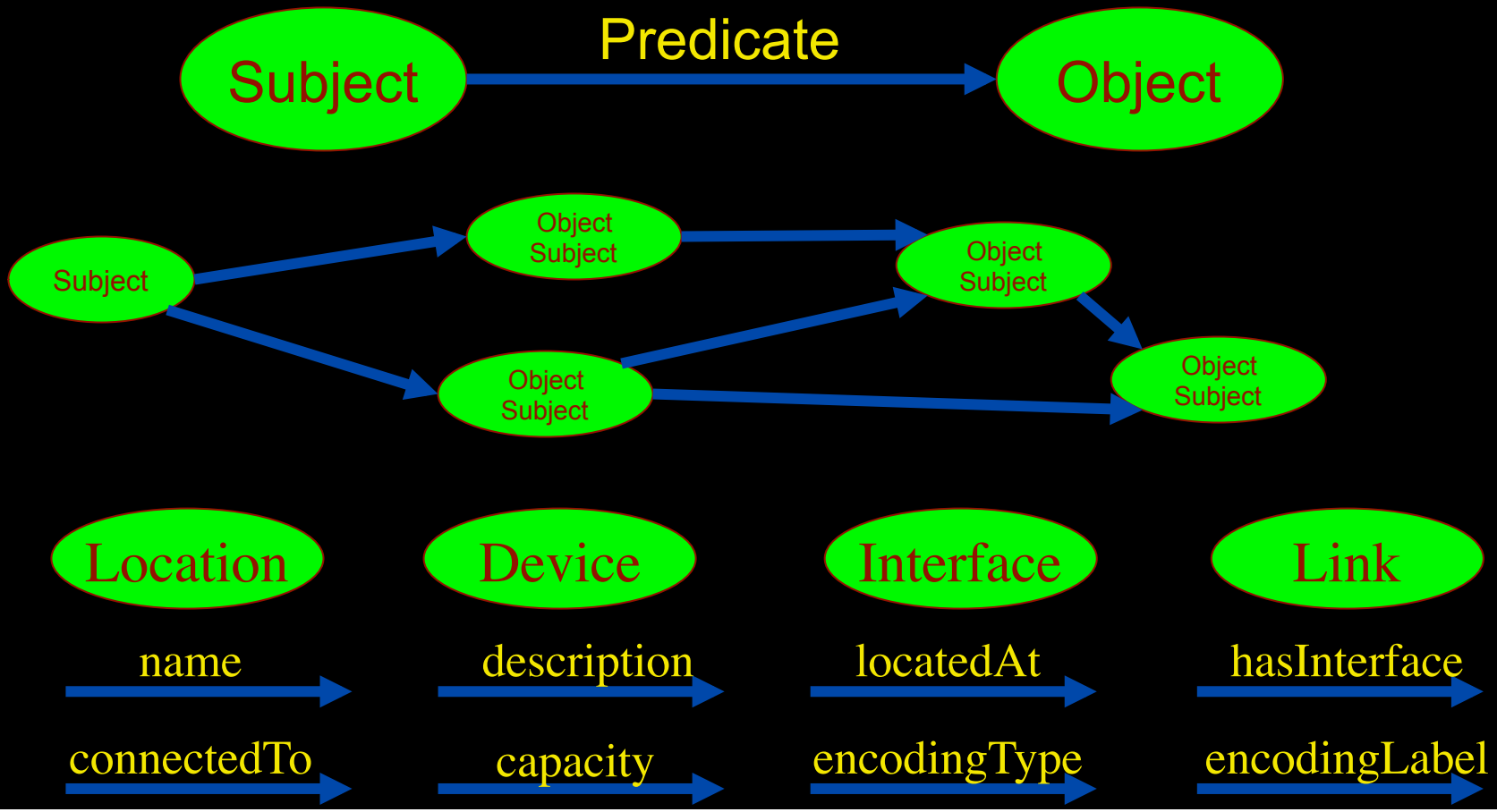
Success depends on the order

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



Network Description Language

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

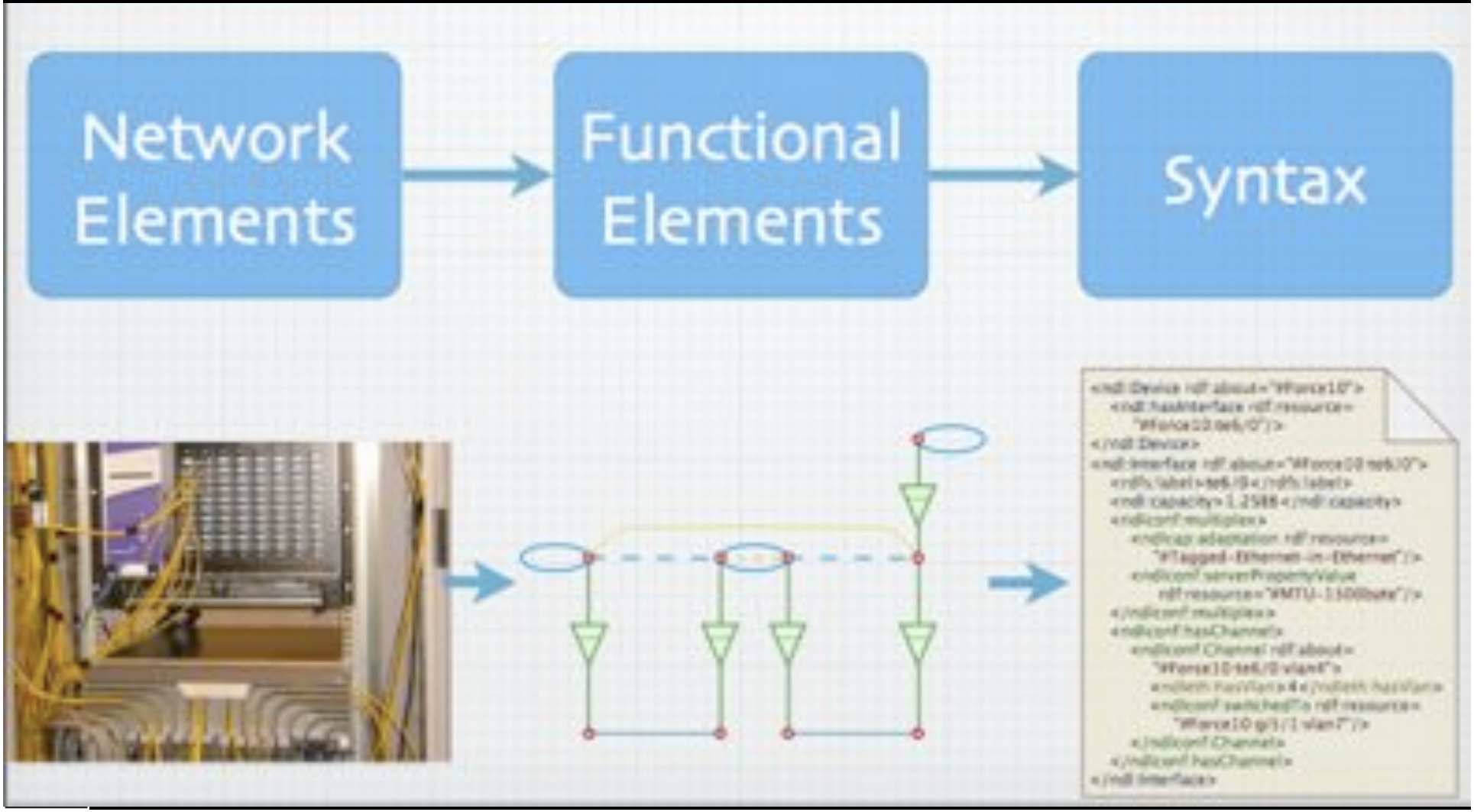


Network Description Language

Choice of RDF instead of XML syntax

Grounded modeling based on G0805 description:

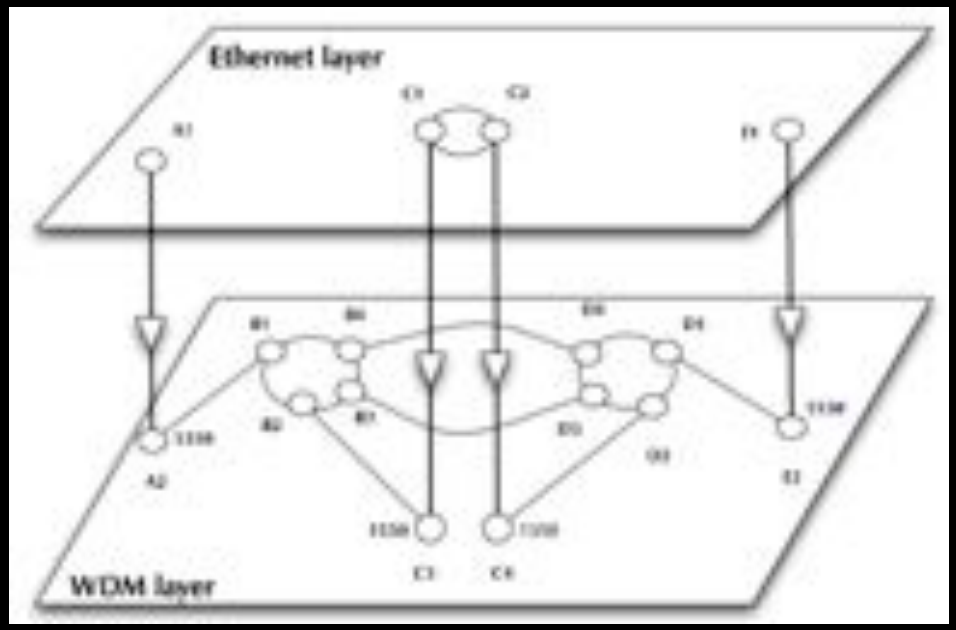
Article: F. Dijkstra, B. Andree, K. Koymans, J. van der Ham, P. Grosso, C. de Laat, *"A Multi-Layer Network Model Based on ITU-T G.805"*

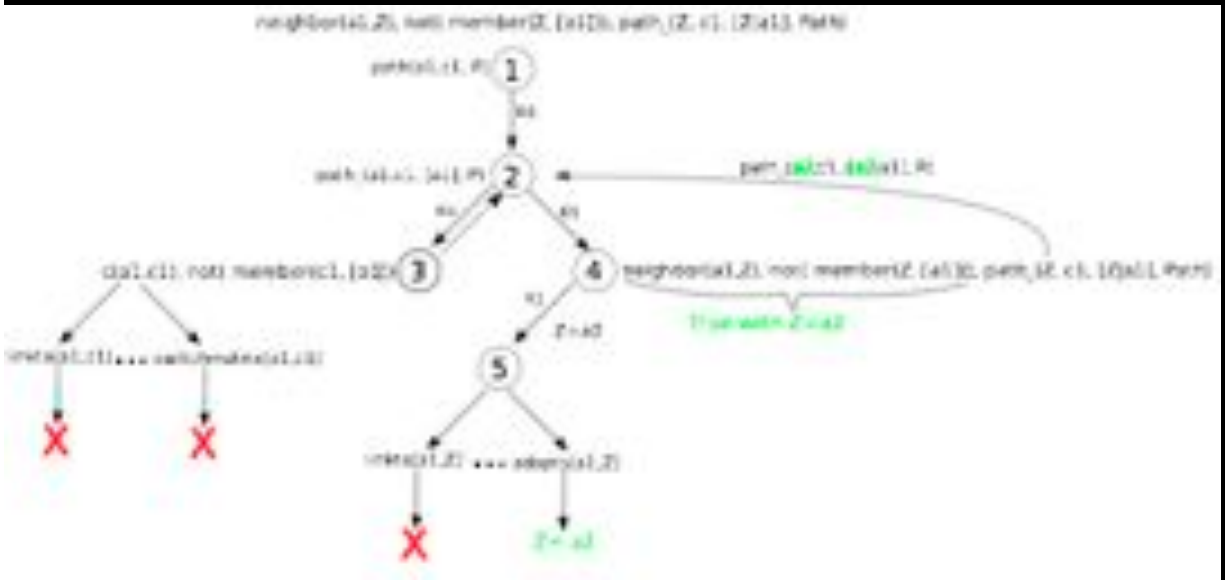


NDL + PROLOG

Research Questions:

- order of requests
- complex requests
- Usable leftovers





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

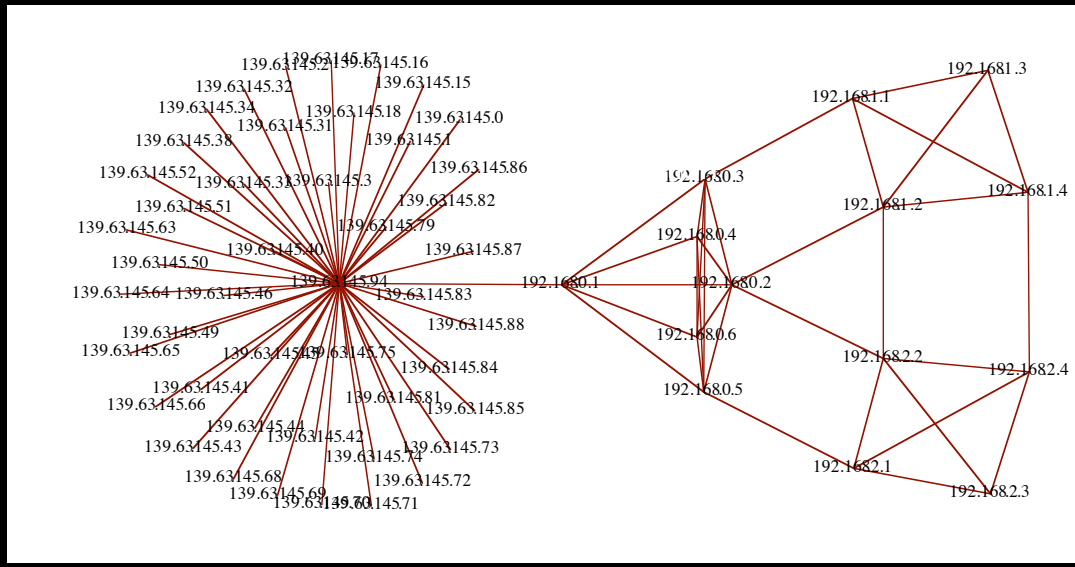
Mathematica enables advanced graph queries, visualizations and real-time network manipulations on UPVNs

Topology matters can be dealt with algorithmically

Results can be persisted using a transaction service built in UPVN

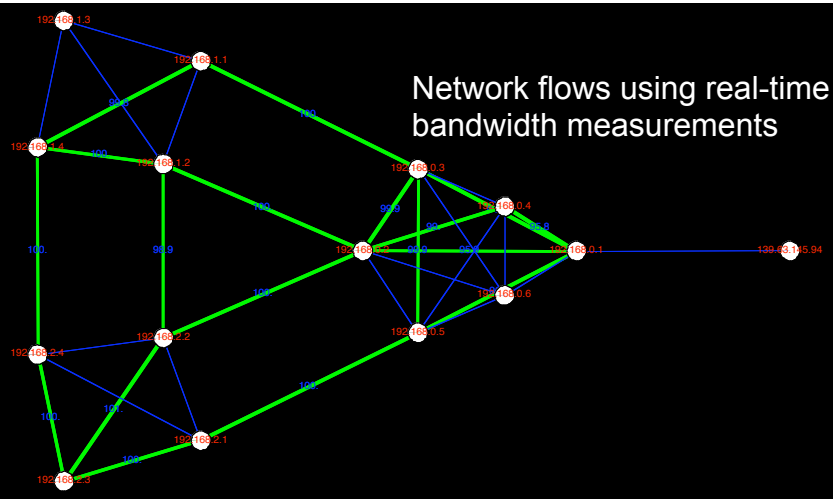
Initialization and BFS discovery of NEs

```
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
Internal links: {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
```



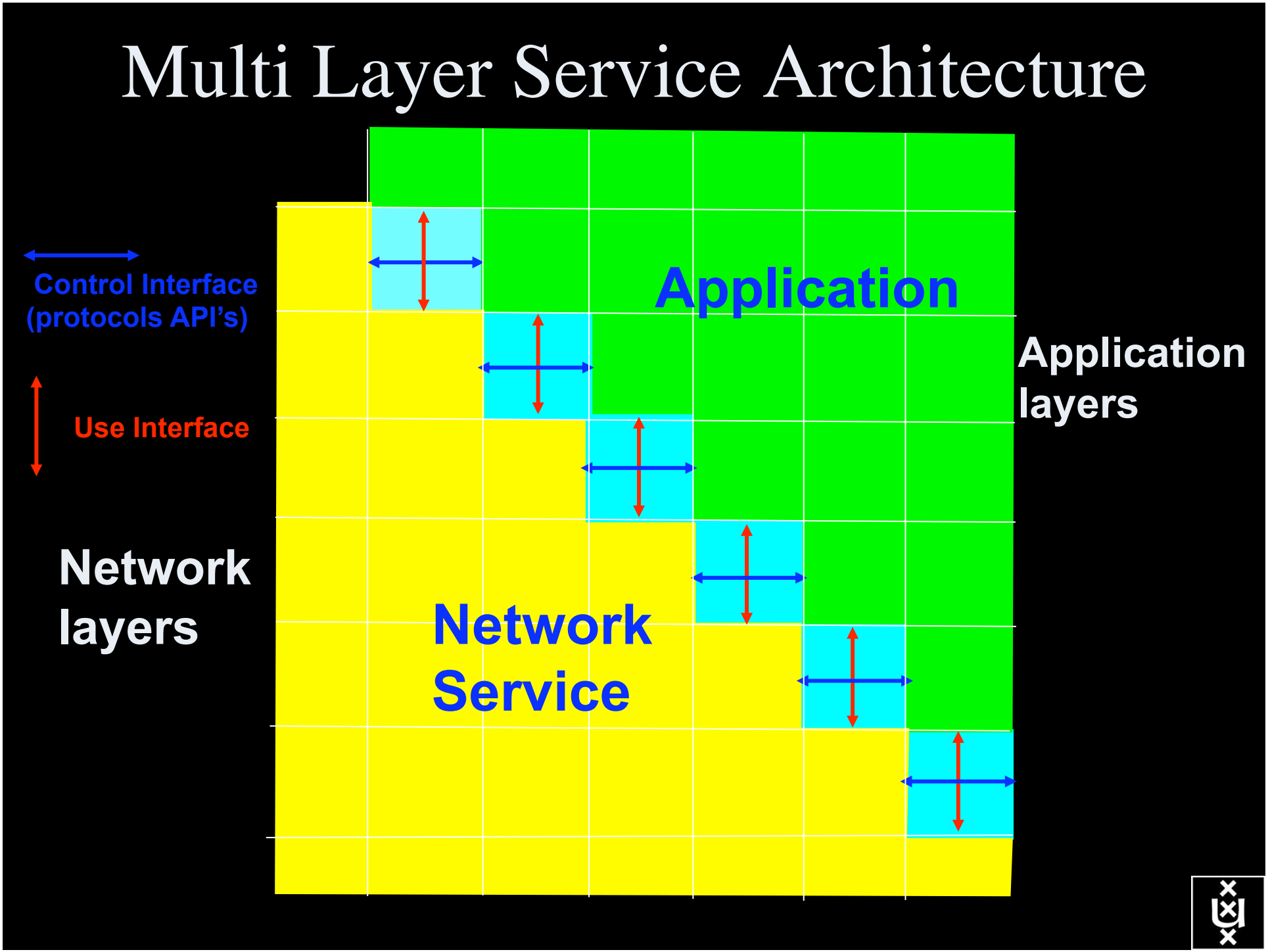
Network flows using real-time bandwidth measurements

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

Interactive programmable networks

Multi Layer Service Architecture



n.a.v. interview met Kees Neggers (SURFnet) & Cees de Laat (UvA)

The COOK Report on Internet Protocol

Technology, Economics, and Policy

©Cook Network Consultants

ICT and E-Science as an Innovation Platform in The Netherlands

A National Research and Innovation Network What Can the US Learn from Dutch Experience?

"The dogmas of the quiet past are inadequate to the stormy present. As our case is new, so we must think anew and act anew." **Abraham Lincoln**

By means of an examination of research networks in Holland, this issue presents some ideas for ways in which an American National Research, Education and Innovation Network could be structured.

possible are carried out by decentralized groups.

Volume XVII, No. 1
February 2009
ISSN 1071-6327

- **BSIK projects GigaPort &**
- **VL-e / e-Science**

SURFNET

cookreport.com

The Basis for a Future Internet?

Optical Hybrid Networks and e-Science as Platforms for Innovation and Tech Transfer

Editor's Note: I continued the discussion begun on No- slide shows our organization within the University and the search department of KPN. He did a lot of virtualization

Questions ?

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

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

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