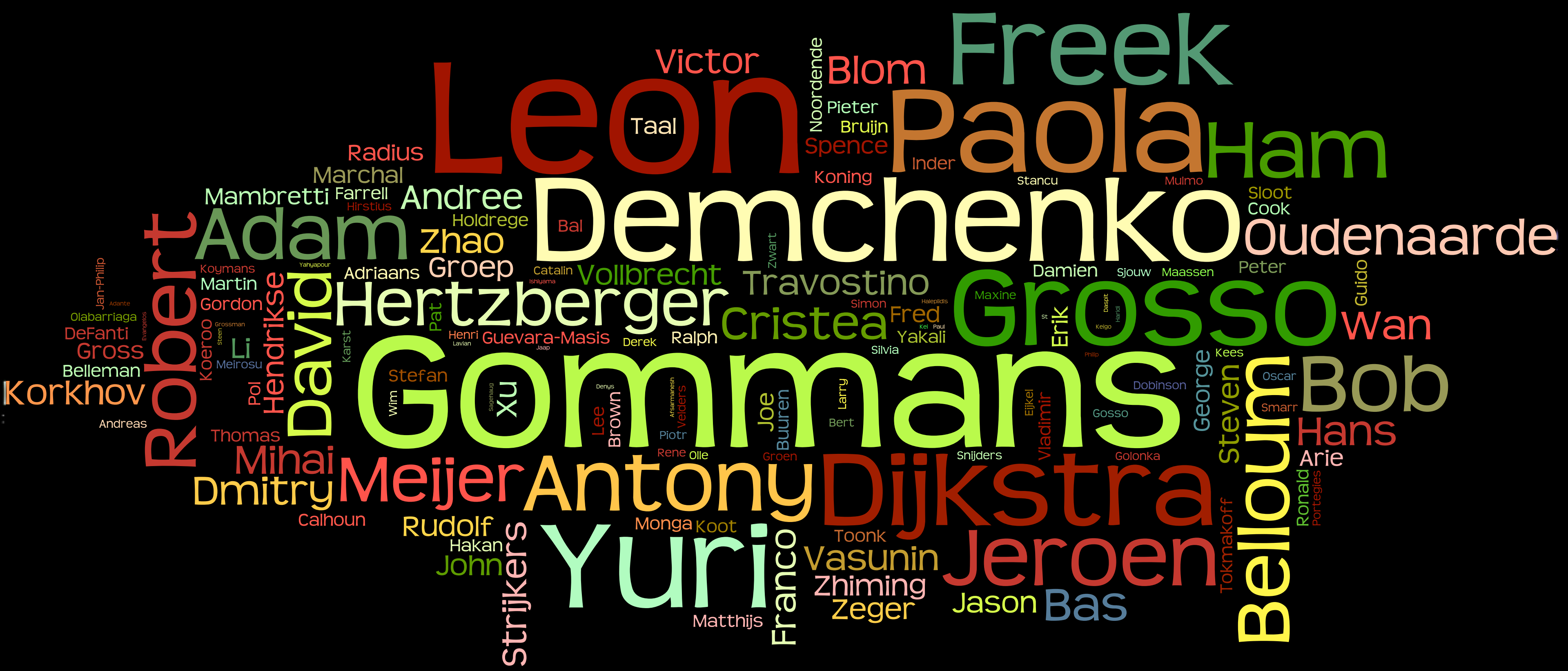
System and Network Engineering Research for Big Data Sciences

Cees de Laat



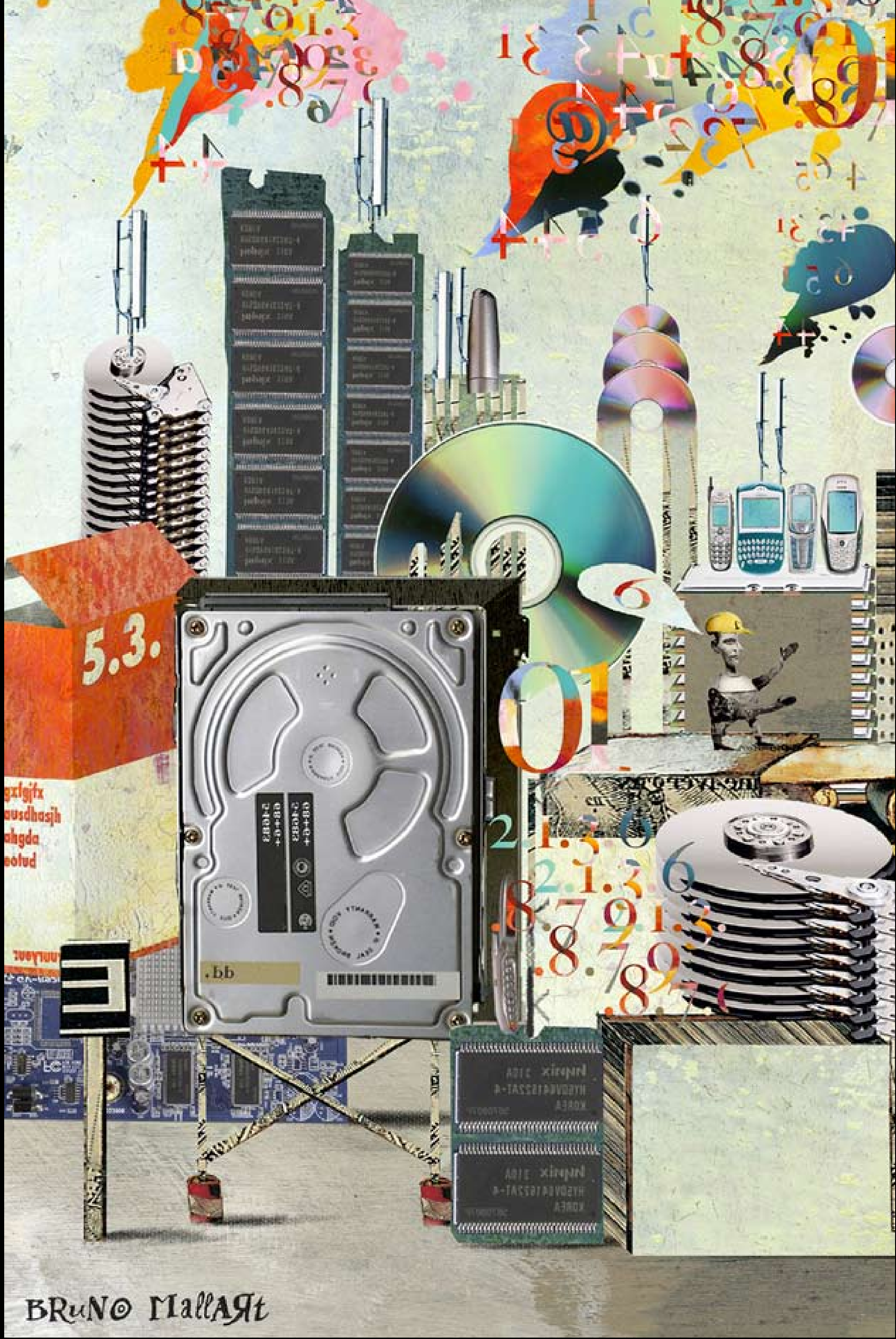
From King's Dutch Academy of Sciences

The Dutch Research Agenda

"Information technology (IT) now permeates all aspects of public, commercial, social, and personal life. bank cards, satnav, and weather radar... IT has become completely indispensable."

"But to guarantee the reliability and quality of constantly bigger and more complicated IT, we will need to find answers to some fundamental questions!"

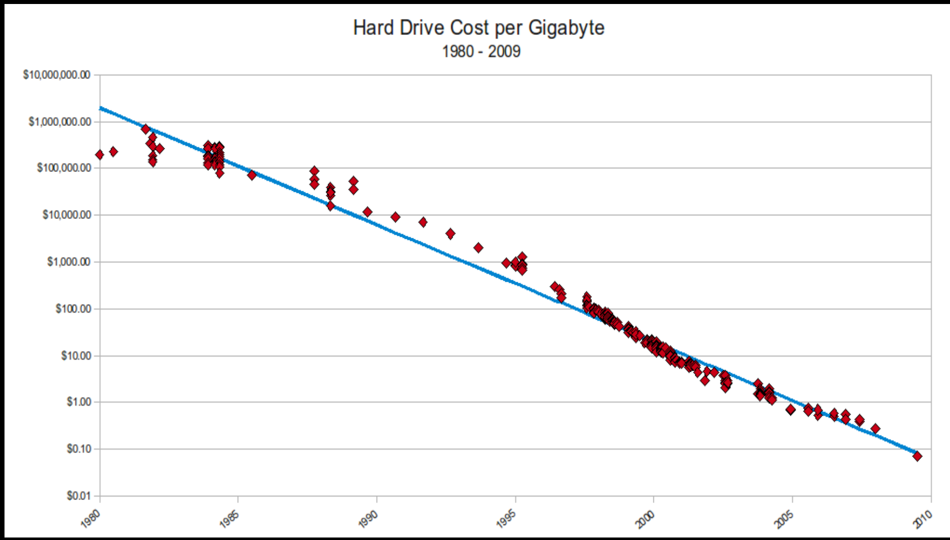
http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf



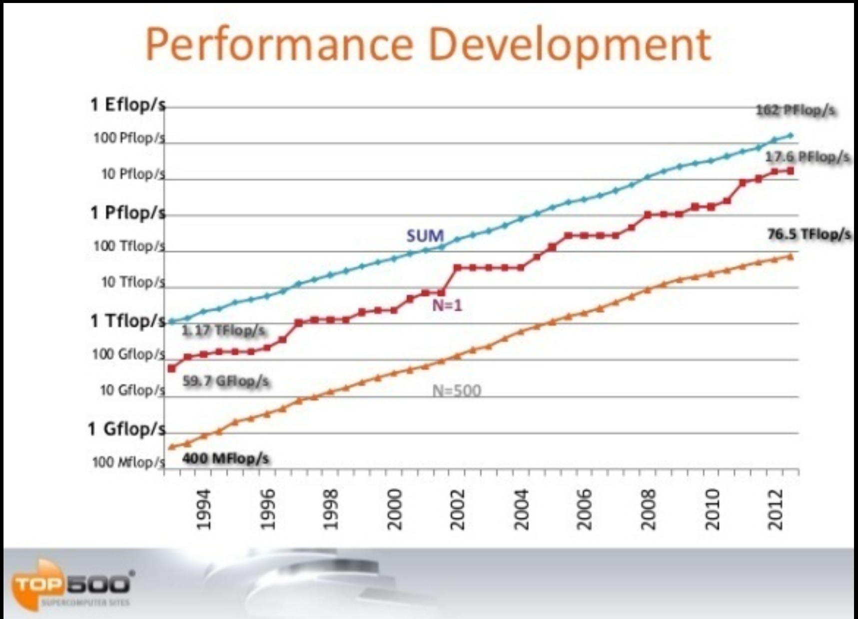
Reliable and Safe!

This omnipresence of IT makes us not only strong but also vulnerable.

- A virus, a hacker, or a system failure can instantly send digital shockwaves around the world.



The hardware and software that allow all our systems to operate is becoming bigger and more complex all the time, and the capacity of networks and data storage is increasing by leaps and bounds.



We will soon reach the limits of what is currently feasible and controllable.

http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf

Reduction of Complexity by Integration

By combining services such as telephony, television, data, and computing capacity within a single network, we can cut down on complexity, energy consumption and maintenance.

- How can we describe and analyze complex information systems effectively?
- How can we specify and measure the quality and reliability of a system?
- How can we combine various different systems?
- How can we design systems in which separate processors can co-operate efficiently via mutual network connections within a much larger whole?
- Can we design information systems that can diagnose their own malfunctions and perhaps even repair them?
- How can we specify, predict, and measure system performance as effectively as possible?

SNE addresses a.o. the **highlighted** questions!

http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf

Mission

Can we create smart data processing infrastructures that are tailored to diverse application needs?

Mission

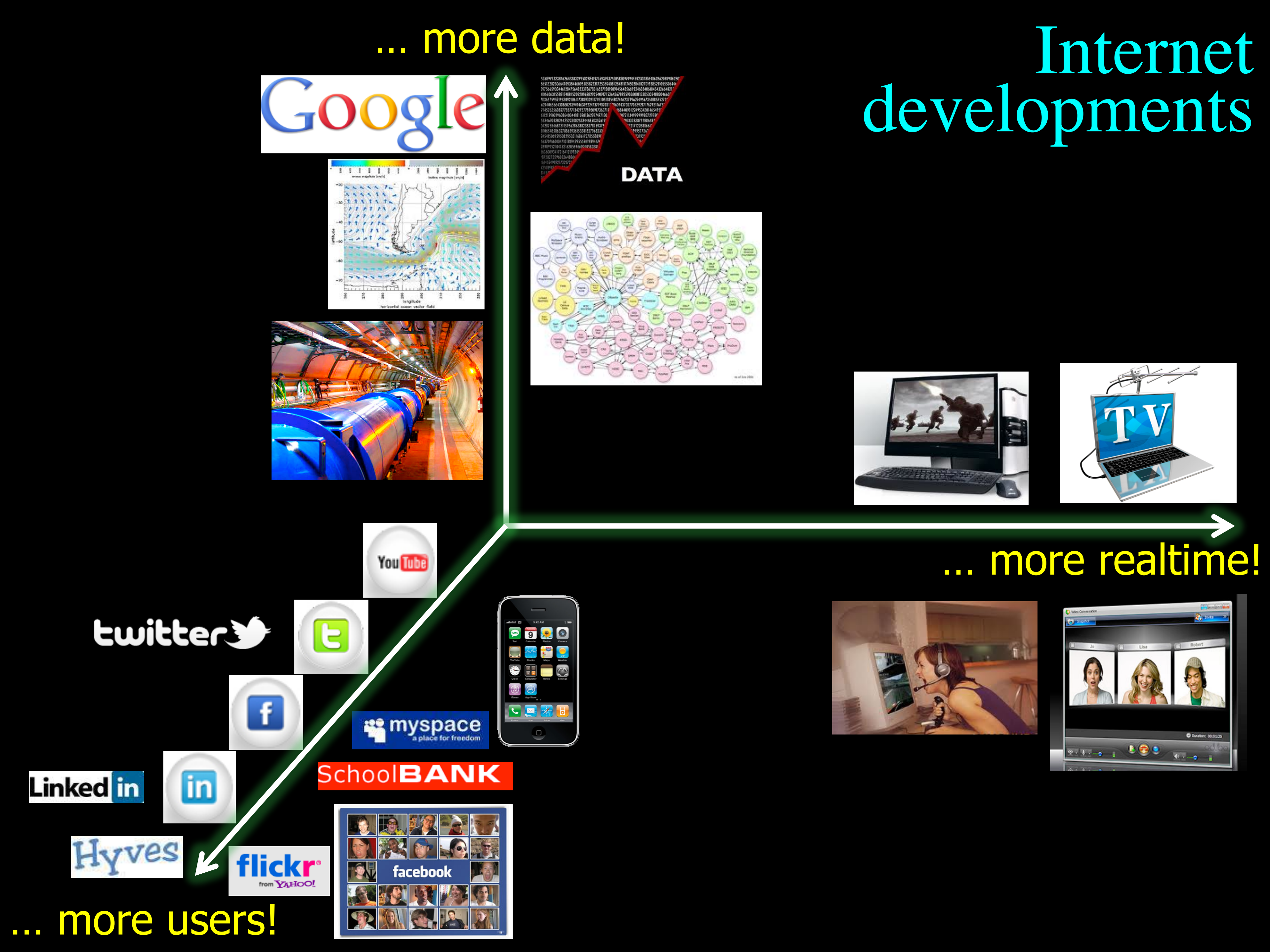
Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

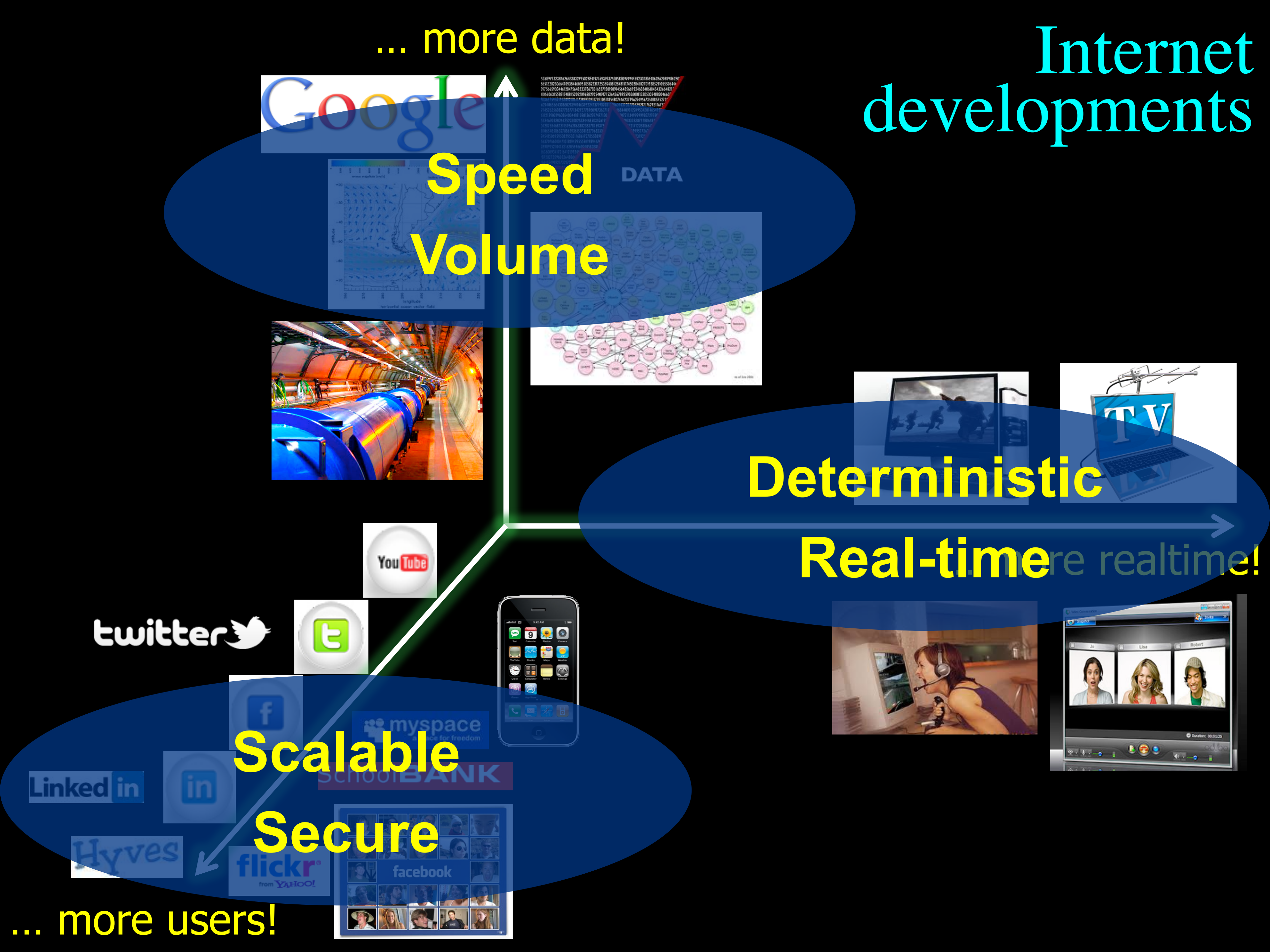
- *Capacity*
- *Capability*
- *Security*
- *Sustainability*
- *Resilience*

Mission

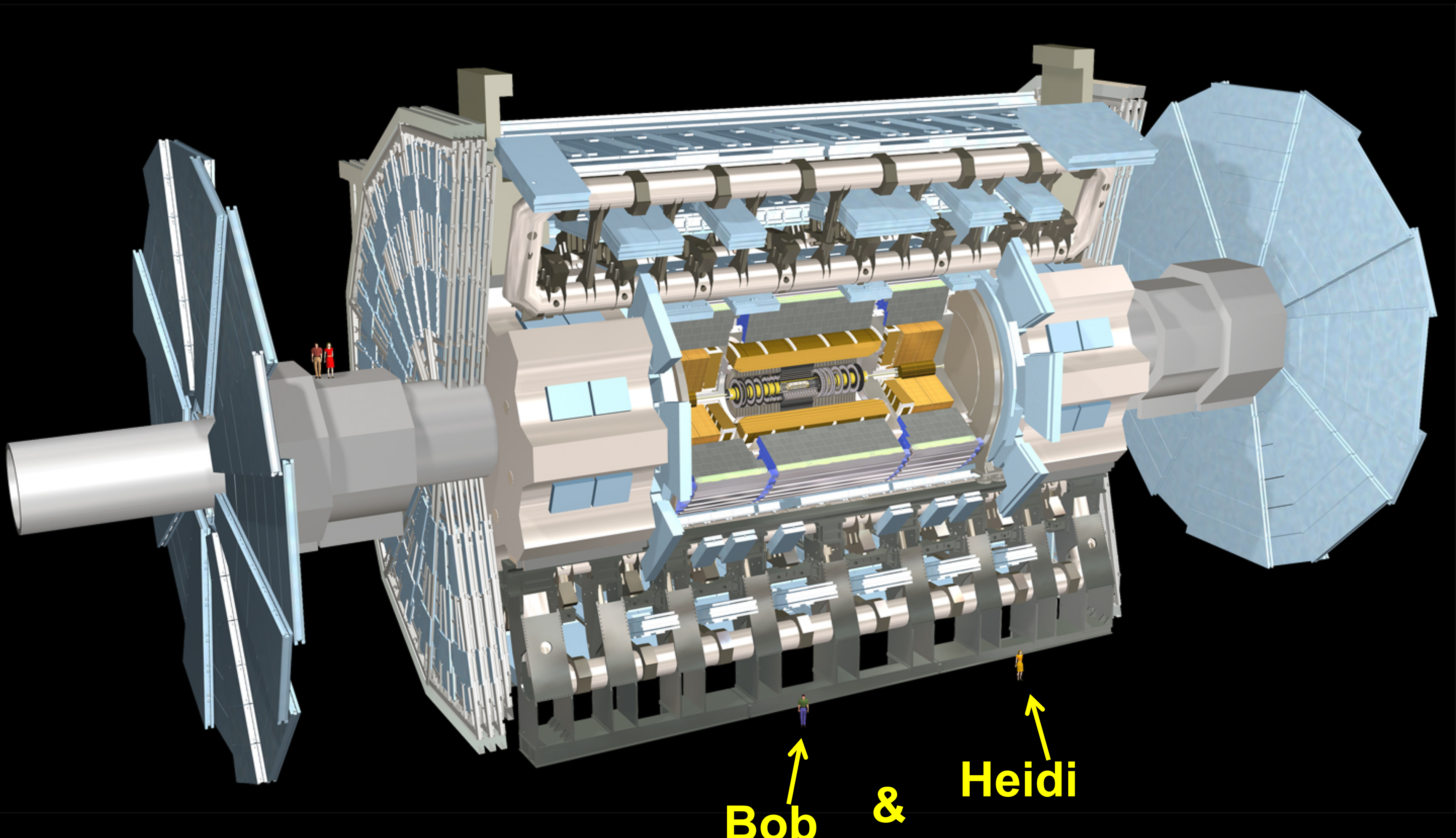
Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Authorization, Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

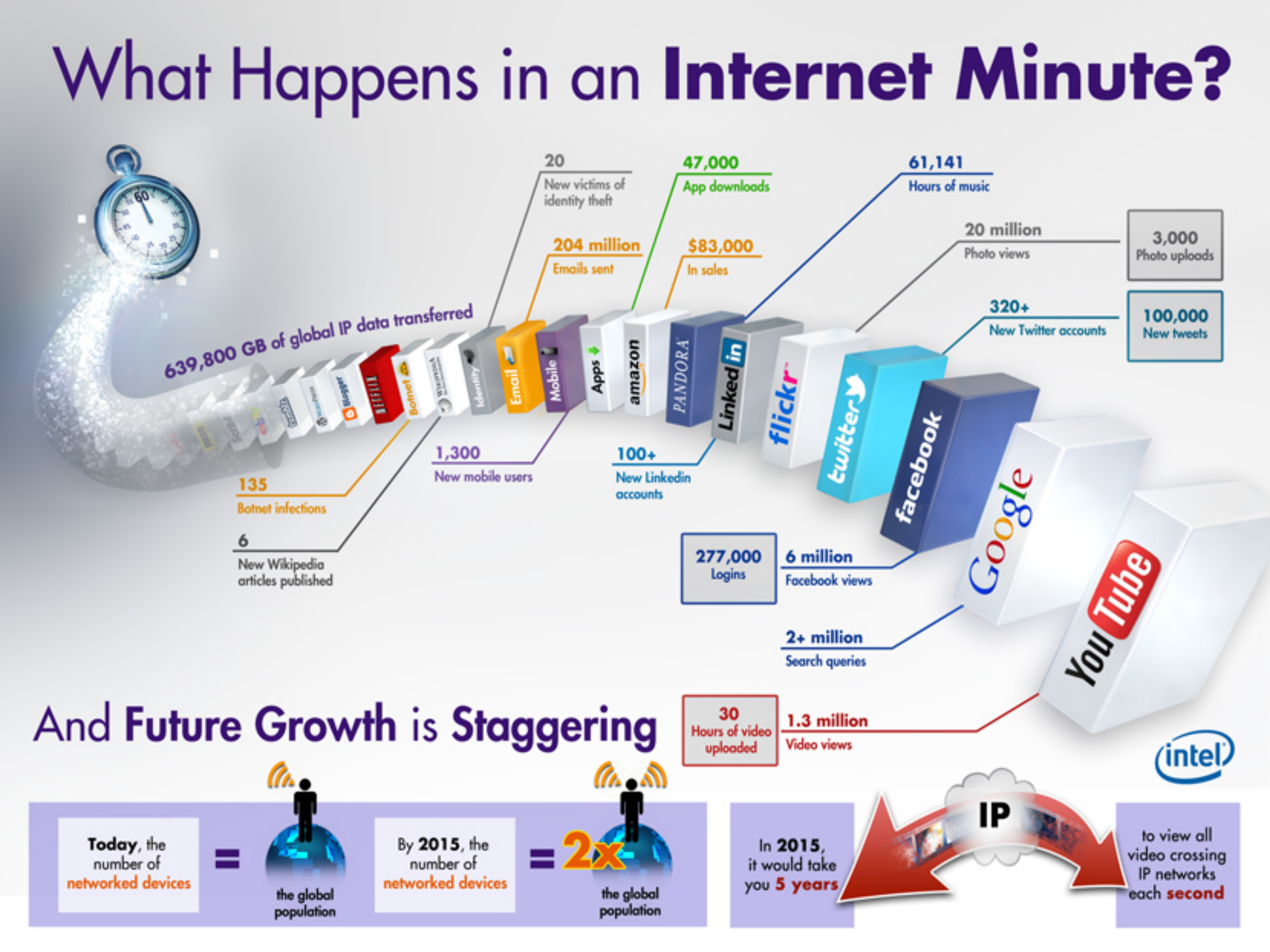

Internet developments
... more data!
Google
DATA
... more realtime!
TV
... more users!
twitter
You Tube
myspace
a place for freedom
SchoolBANK
Linked in
Hyves
flickr
from YAHOO!
facebook

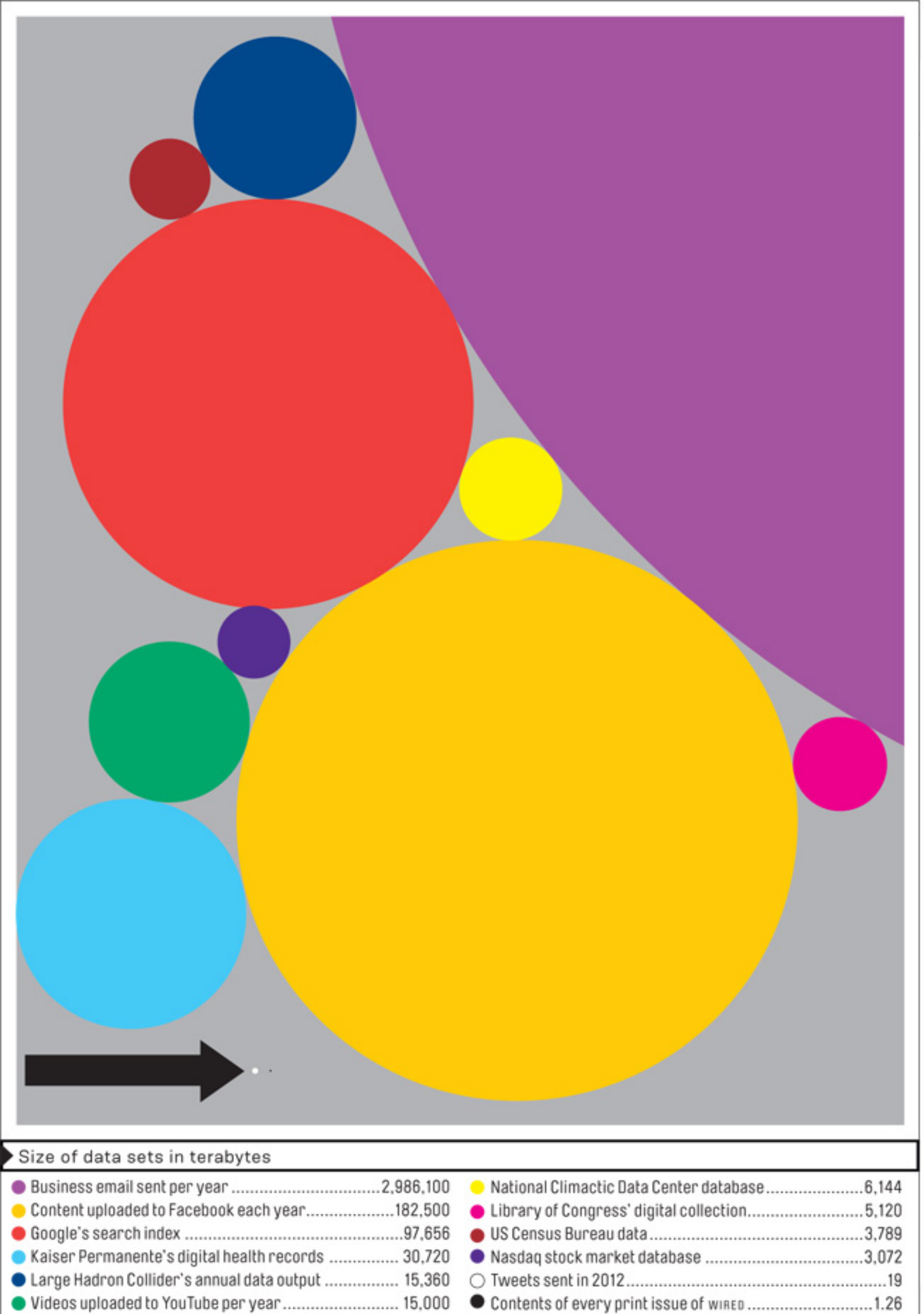

Internet developments
... more data!
Speed
Volume
DATA
Deterministic
Real-time
re realtime!
Scalable
Secure
... more users!
Google
You Tube
twitter
myspace
a place for freedom
schoolBANK
Linked in
Hyves
flickr
from YAHOO!
facebook
TV

ATLAS detector @ CERN Geneve



ATLAS detector @ CERN Geneve

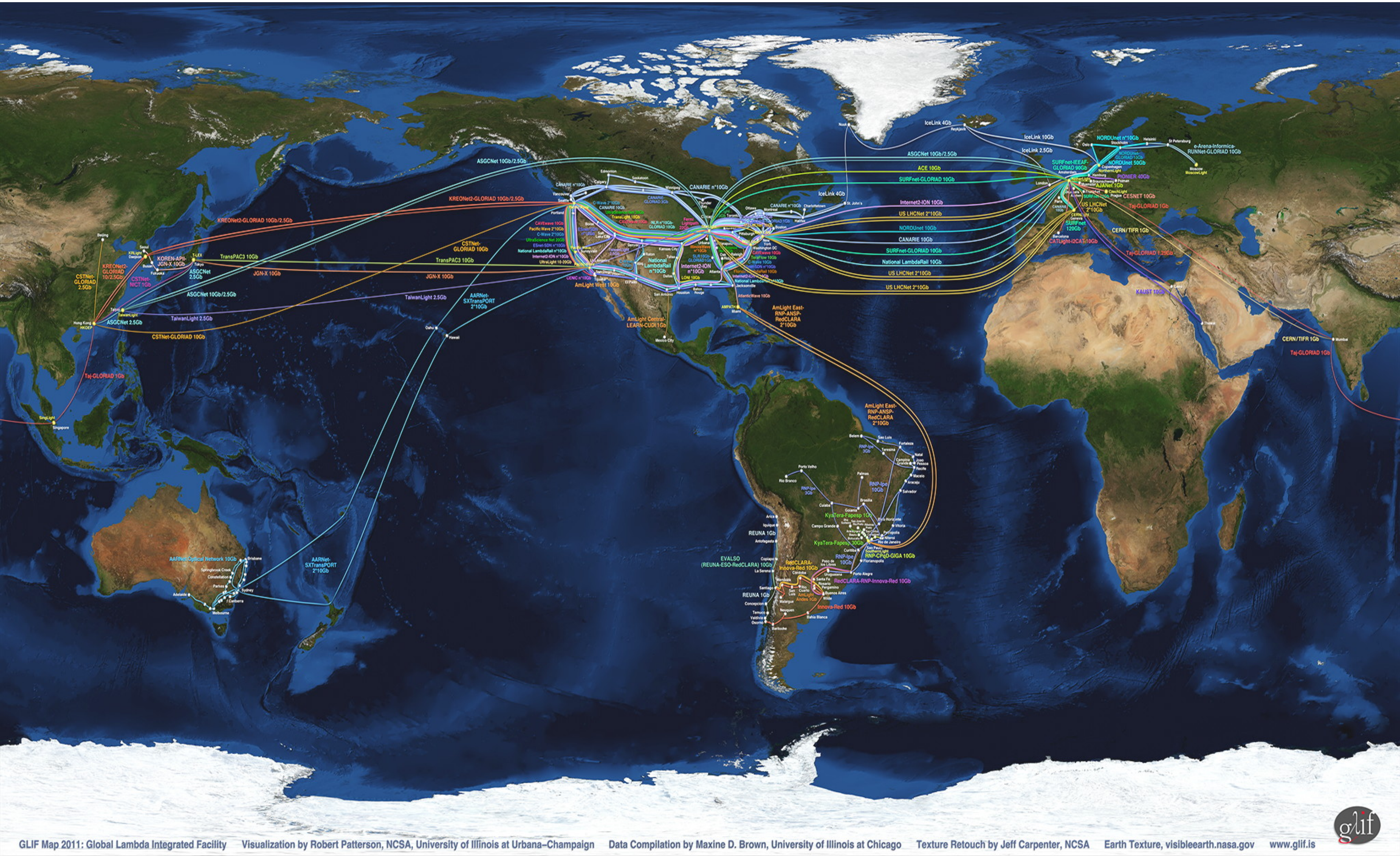

What Happens in an Internet Minute?
639,800 GB of global IP data transferred
20
New victims of identity theft
204 million
Emails sent
47,000
App downloads
$83,000
In sales
61,141
Hours of music
20 million
Photo views
3,000
Photo uploads
320+
New Twitter accounts
100,000
New tweets
135
Botnet infections
6
New Wikipedia articles published
1,300
New mobile users
100+
New Linkedin accounts
277,000
Logins
6 million
Facebook views
2+ million
Search queries
30
Hours of video uploaded
1.3 million
Video views
Botnet
Identity
Email
Mobile
Apps
amazon
PANDORA
LinkedIn
flickr
twitter
facebook
Google
YouTube
And Future Growth is Staggering
Today, the number of networked devices = the global population
By 2015, the number of networked devices = 2x the global population
In 2015, it would take you 5 years to view all video crossing IP networks each second
IP
intel



There is always a bigger fish

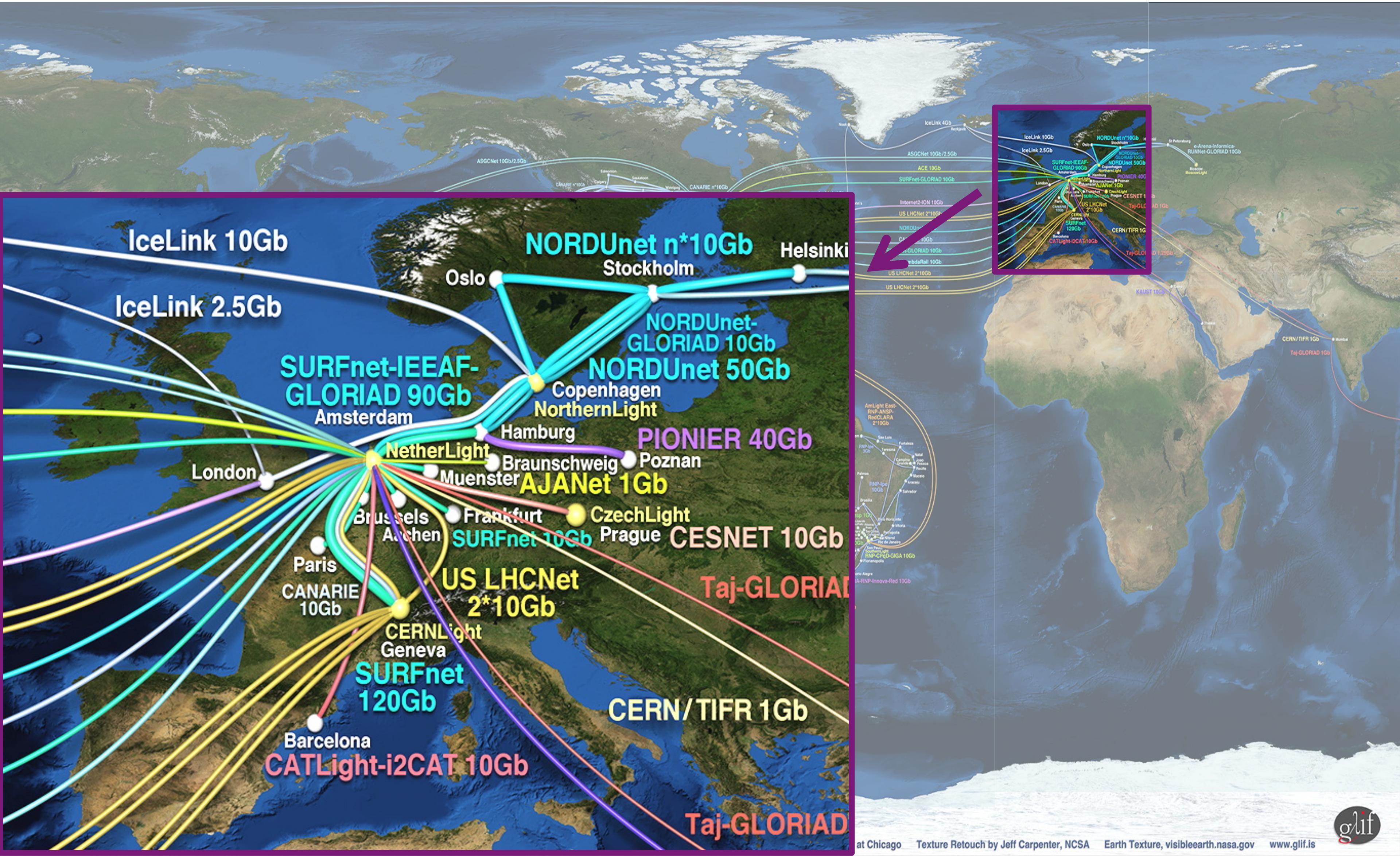
The GLIF – LightPaths around the World

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.



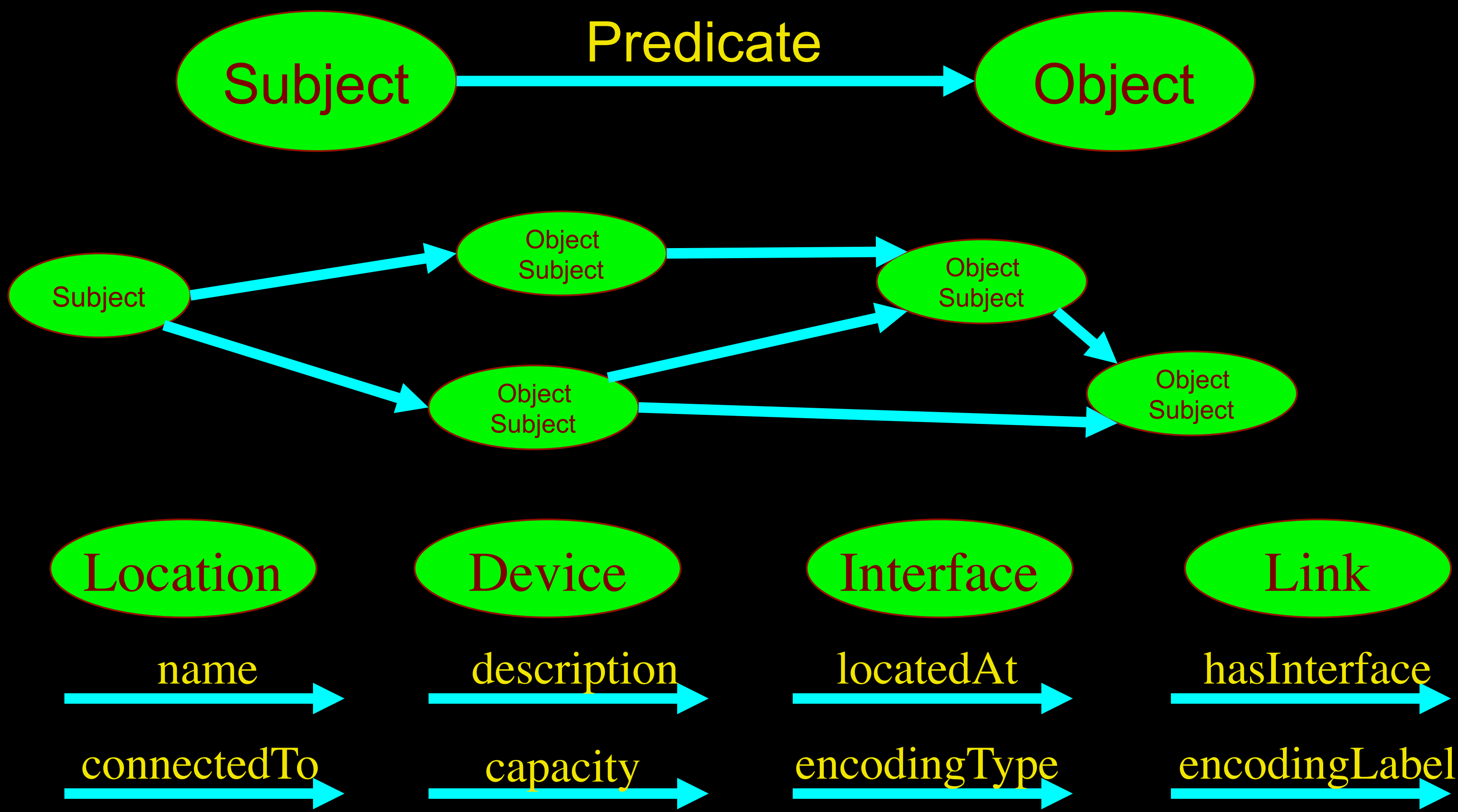
The GLIF – LightPaths around the World

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.

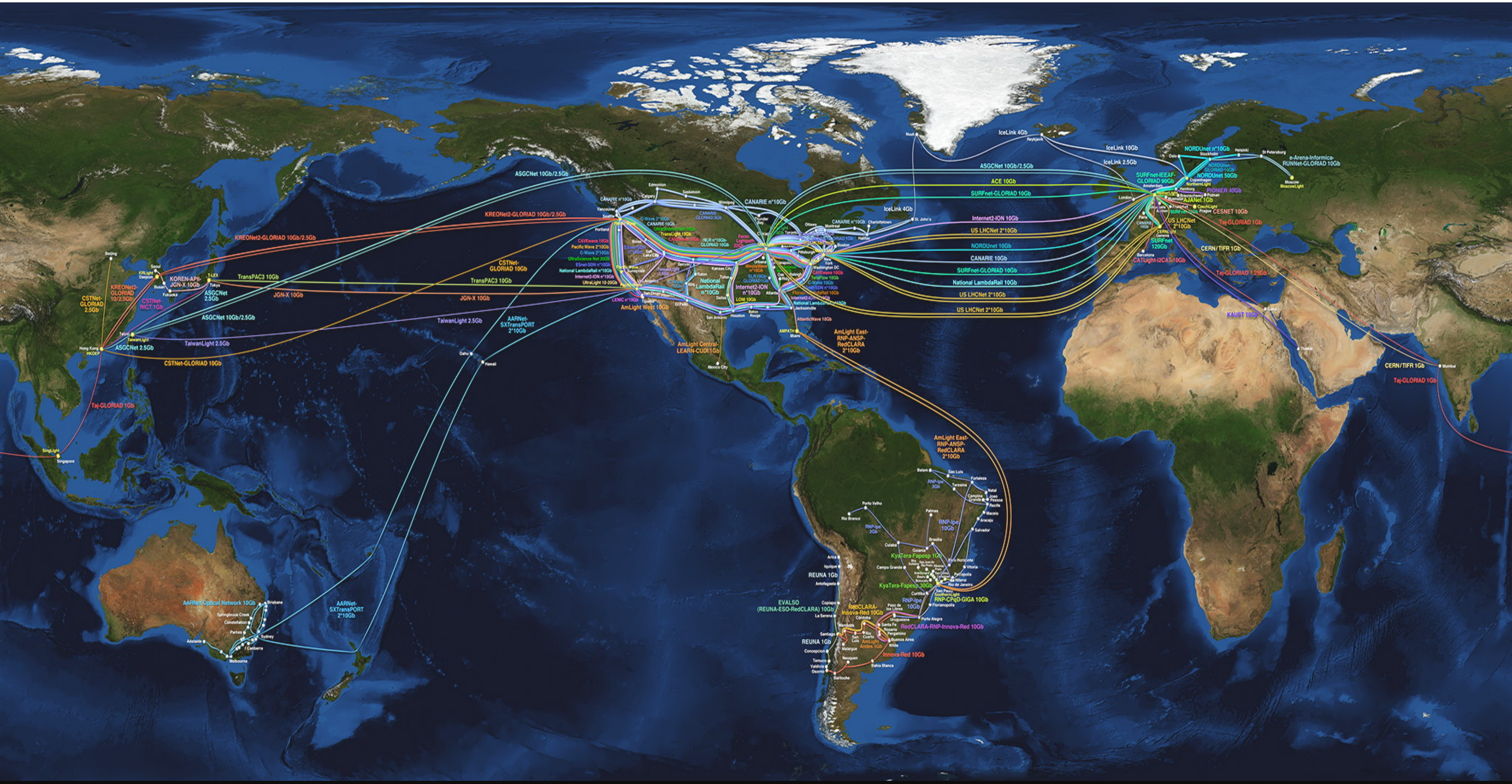


LinkedIN for Infrastructure

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



The GLIF – LightPaths around the World

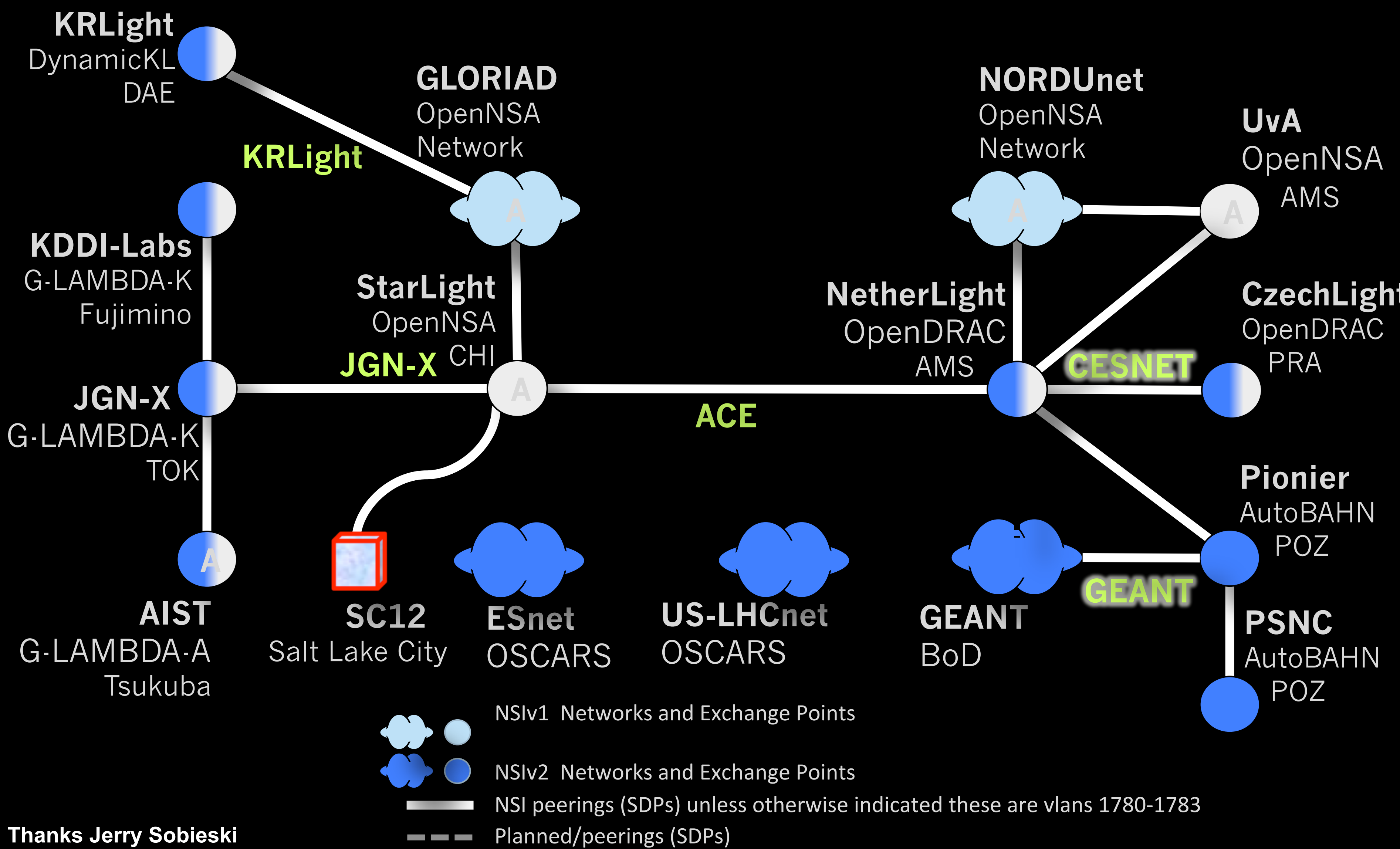


We investigate: TOMTOM for complex networks!

Automated GOLE + NSI

Joint NSI v1+v2 Beta Test Fabric   Nov 2012

Ethernet Transport Service

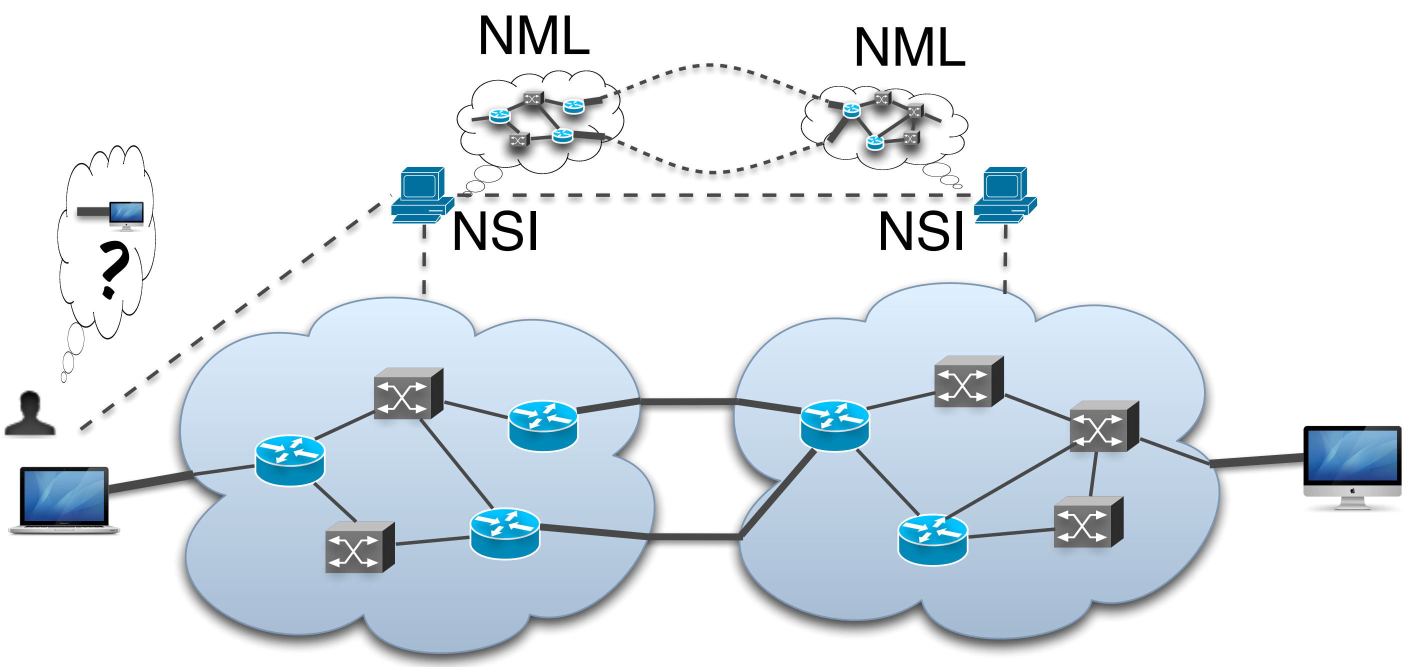


Thanks Jerry Sobieski

Network Topology Description

Network topology research supporting automatic network provisioning

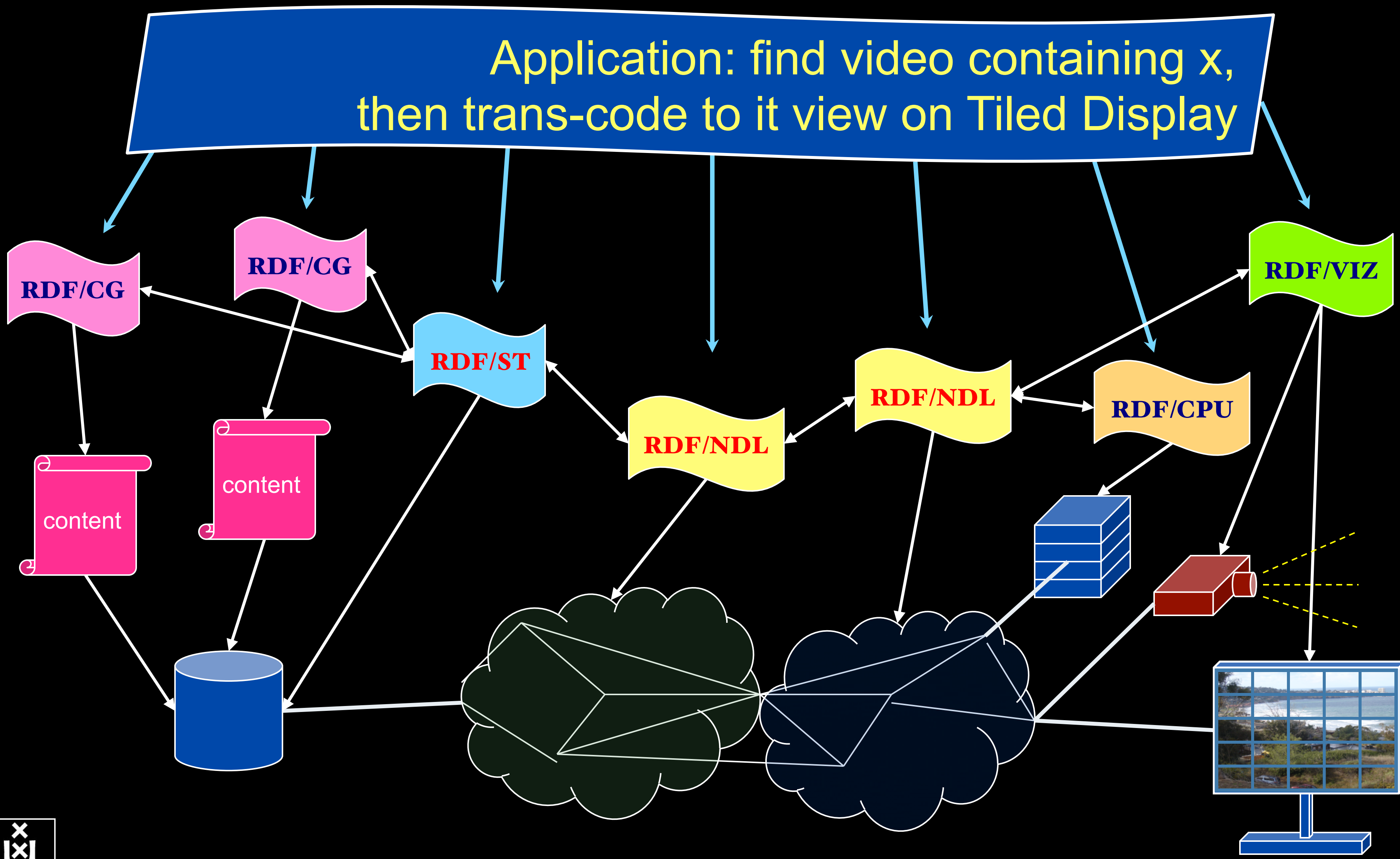
- Inter-domain networks
- Multiple technologies
- Based on incomplete information
- Possibly linked to other resources



http://redmine.ogf.org/projects/nml-wg
http://redmine.ogf.org/projects/nsi-wg

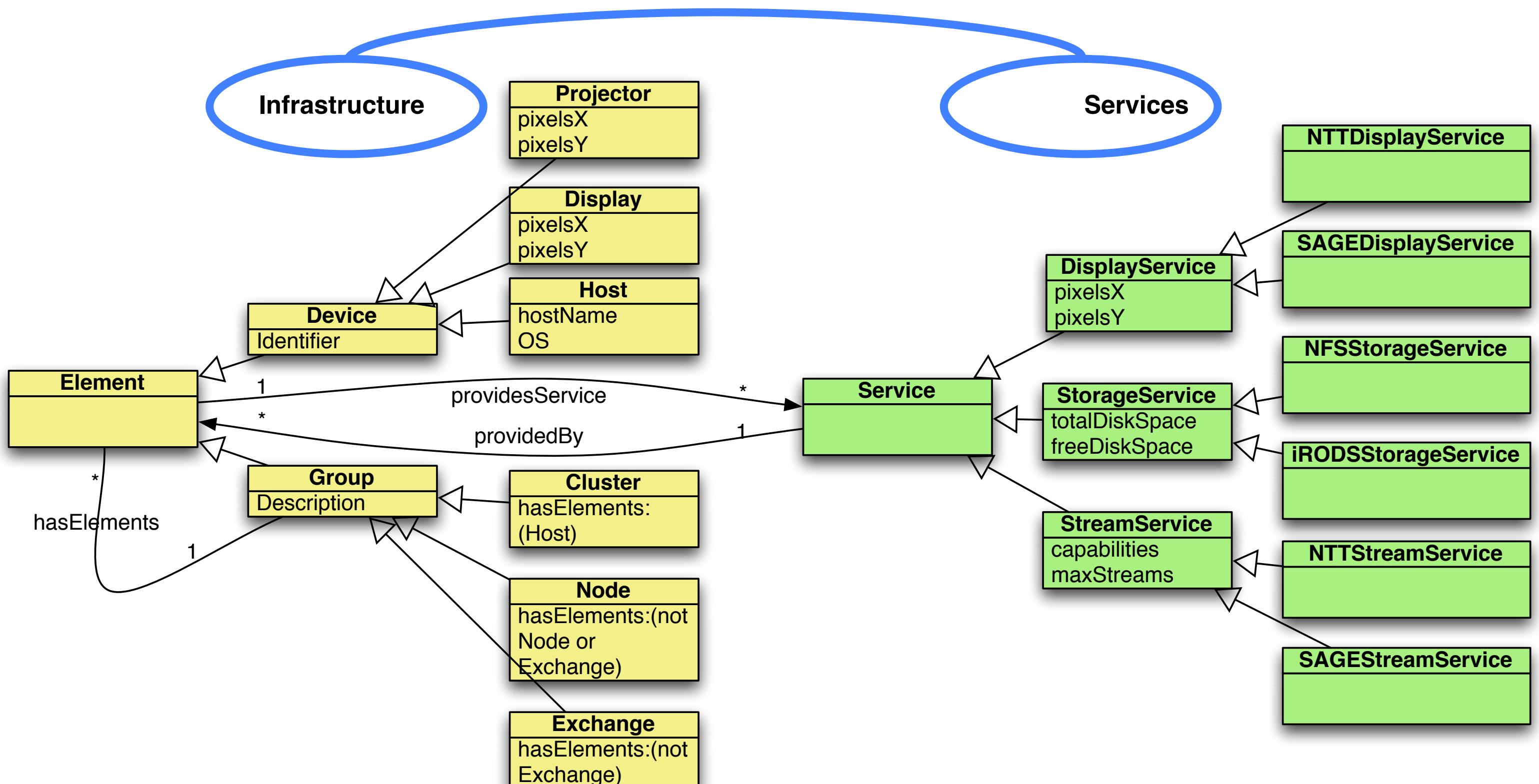
http://sne.science.uva.nl/ndl

RDF describing Infrastructure
“I want”



Information Modeling

Define a common information model for ***infrastructures*** and ***services***.
Base it on Semantic Web.



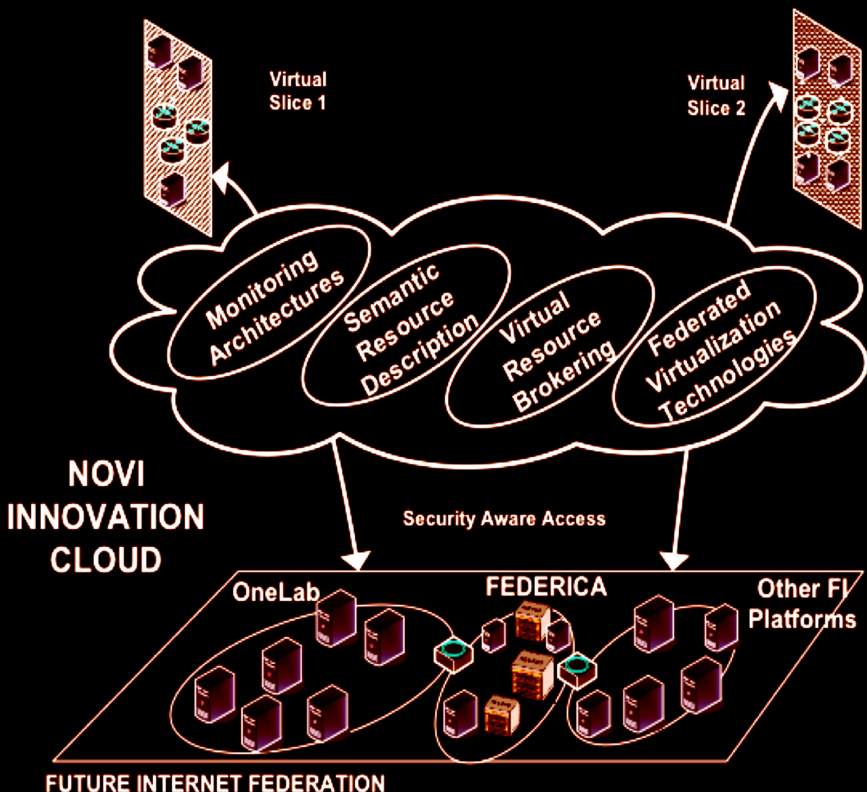
J. van der Ham, F. Dijkstra, P. Grosso, R. van der Pol, A. Toonk, C. de Laat
A distributed topology information system for optical networks based on the semantic web, Elsevier Journal on Optical Switching and Networking, Volume 5, Issues 2-3, June 2008, Pages 85-93

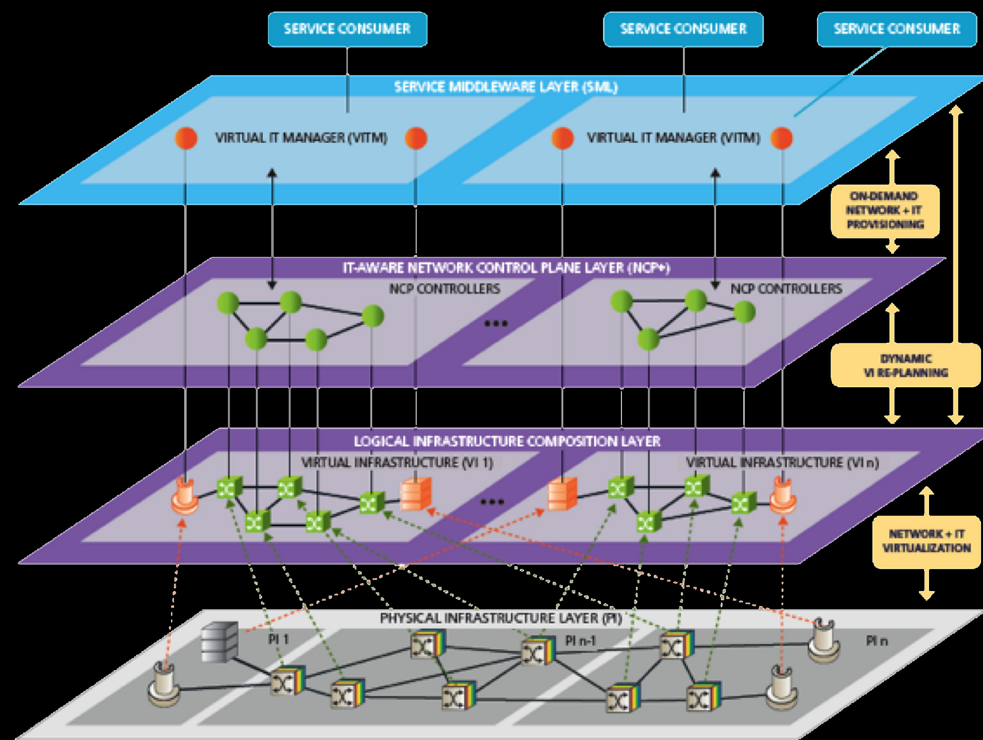
R.Koning, P.Grosso and C.de Laat
Using ontologies for resource description in the CineGrid Exchange
In: Future Generation Computer Systems (2010)

NOVI: Networking Innovations Over Virtualized Infrastructures.
GEYSERS: Generalized Architecture for Dynamic Infrastructure Services.

You can gain more by take all pieces of the puzzle into account.

1. describe ***all*** resources (networks, computing and storage facilities)
2. *optimize the computing problem*, instead of only its network aspect



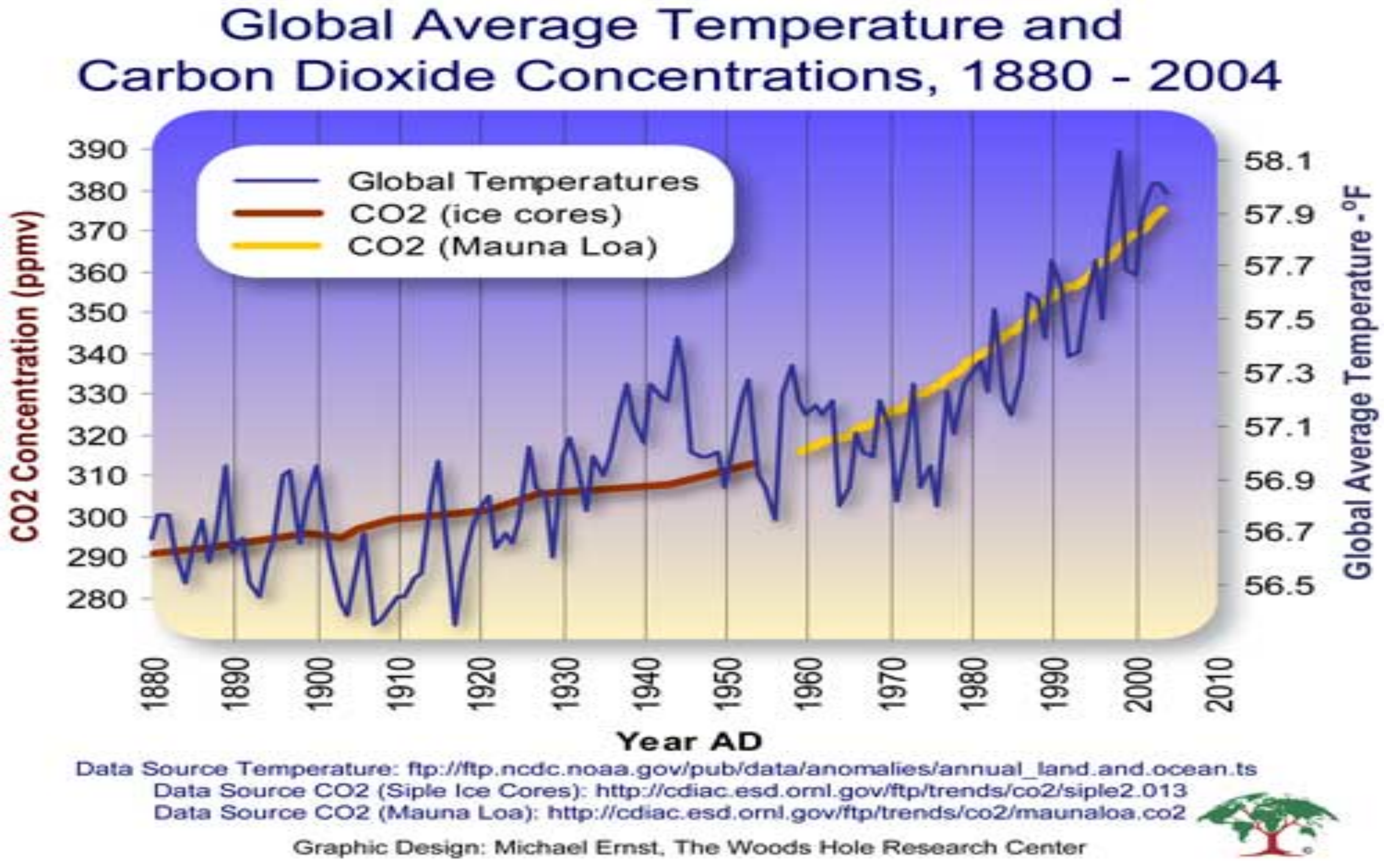


L. Lymberopoulos , P. Grosso, D. Kalogeras , C.Papagianni , C. de Laat , and V. Maglaris, “Ontology-based Policy Based Management for Federated Virtualised Platforms”, Third IFIP/IEEE International Workshop on Management of the Future Internet - May 2011.

J. van der Ham, C. Papagianni, J. Steger, P.Matray, J. Kryftis, P. Grosso and L. Lymberopoulos, “Challenges of an Information Model for Federating Virtualized Infrastructures”, 5th Intl. DMTF Academic Alliance Workshop on Systems and Virtualization Management: Standards and the Cloud, Paris 24 Oct. 2011.

Demchenko, Y., C. Ngo, C. de Laat, T. Wlodarczyk, C. Rong, W. Ziegler, Security Infrastructure for On-demand Provisioned Cloud Infrastructure Services, Proc. 3rd IEEE Conf. on Cloud Computing Technologies and Science (CloudCom2011), 29 November - 1 December 2011, Athens, Greece. (Best Paper Award)

Need for GreenIT



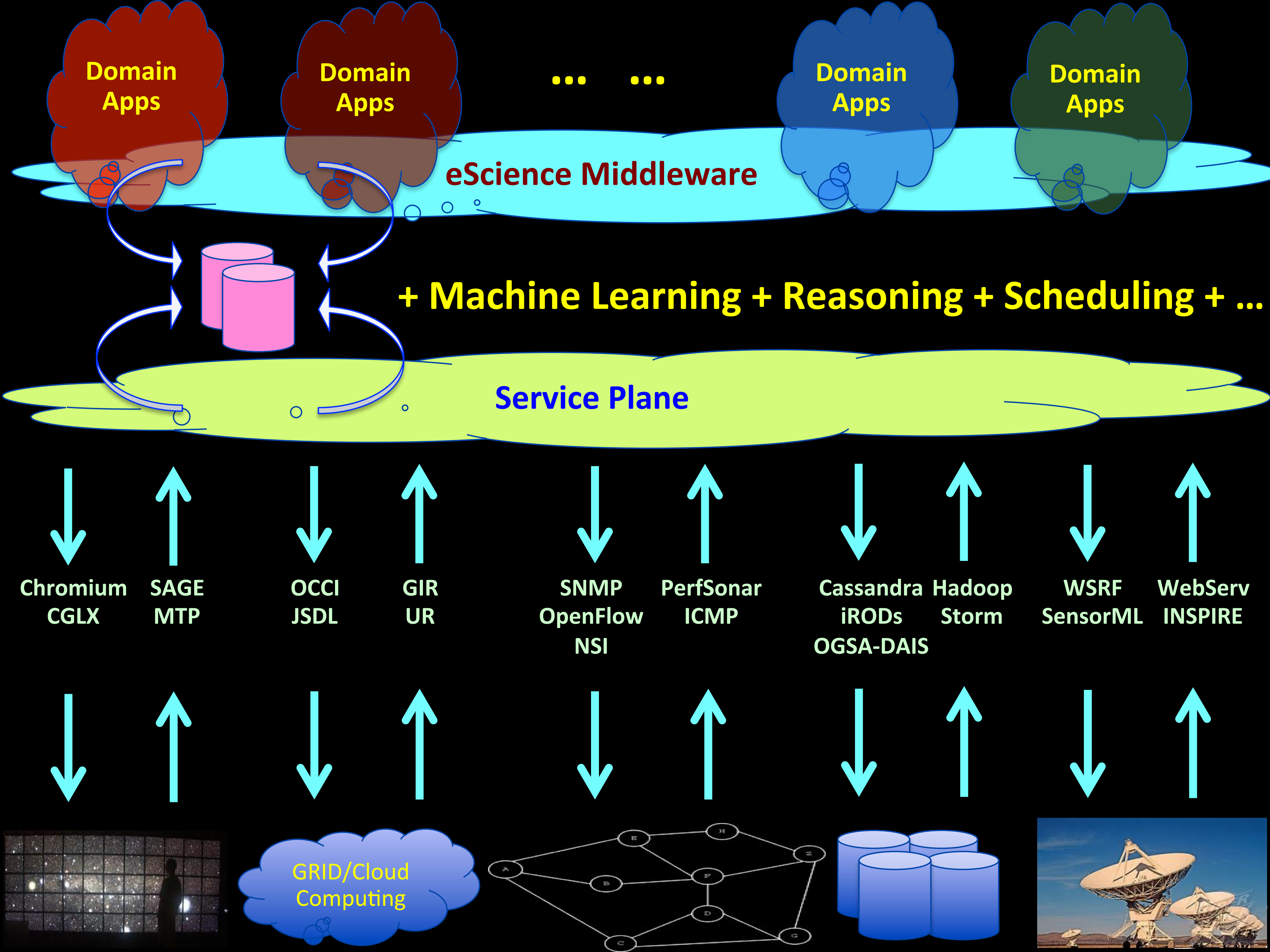
Greening the Processing System

ECO-Scheduling

I want to

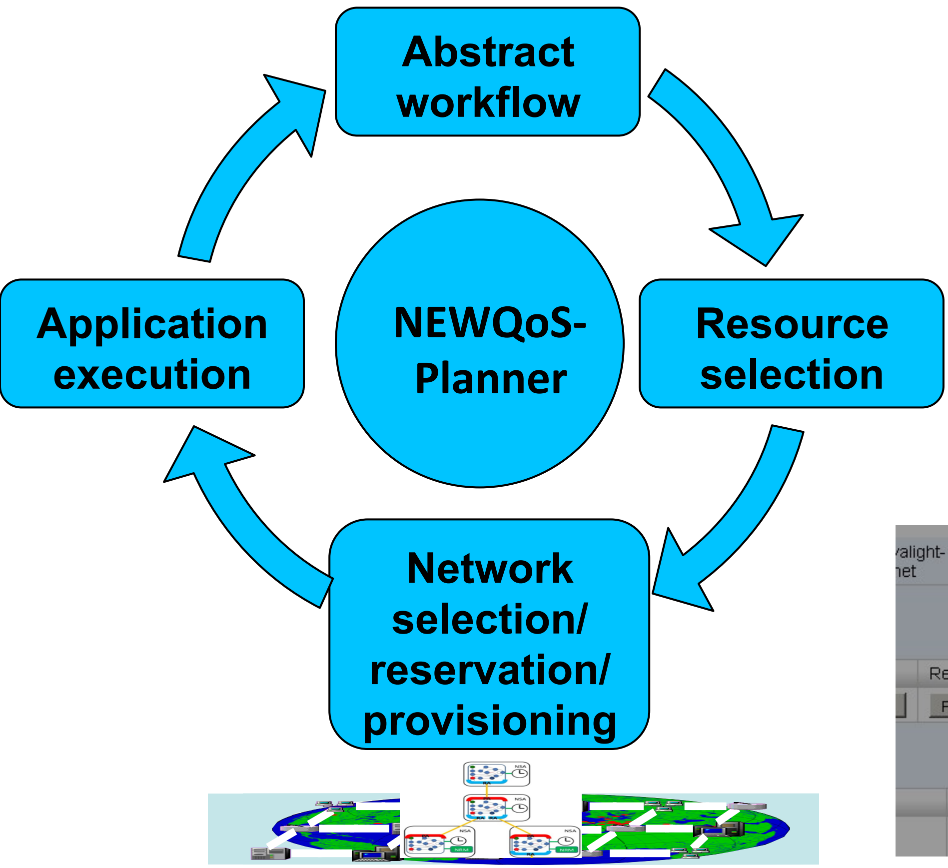
"Show Big Bug Bunny in 4K on my Tiled Display using green Infrastructure"

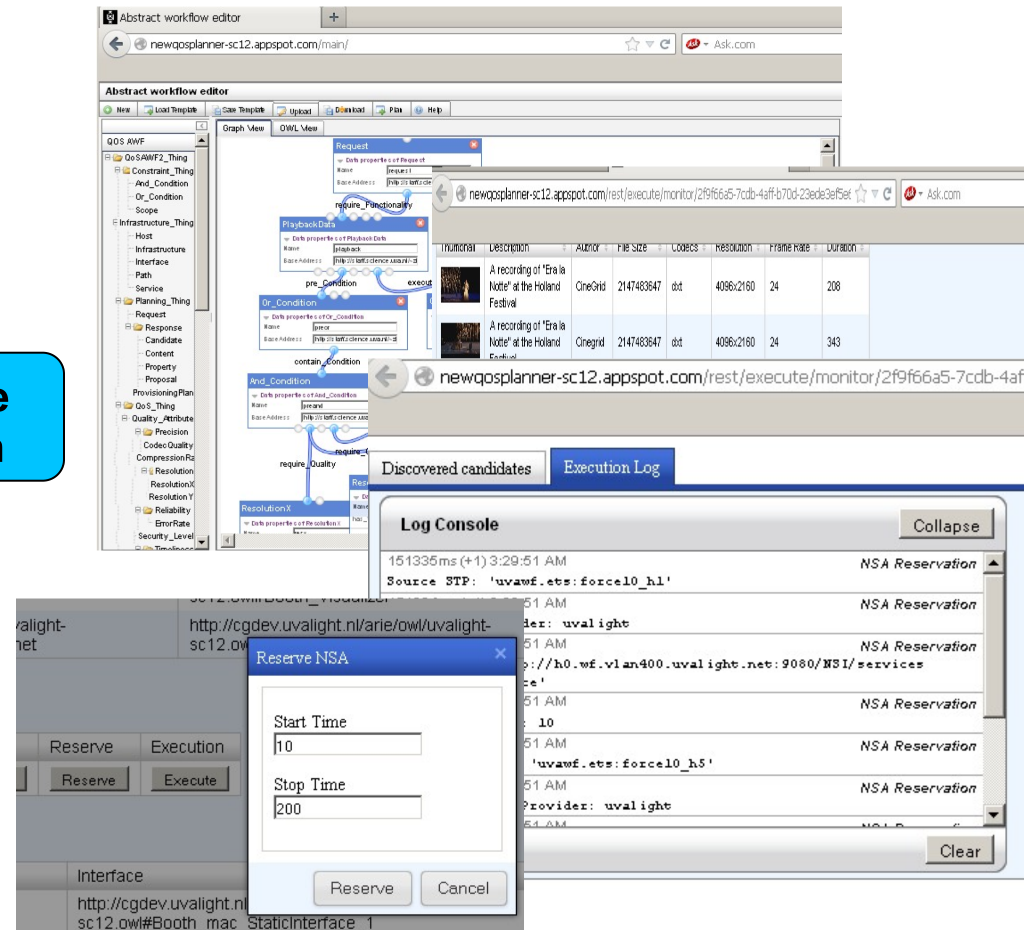
- Big Bugs Bunny can be on multiple servers on the Internet.
- Movie may need processing / recoding to get to 4K for Tiled Display.
- Needs deterministic Green infrastructure for Quality of Experience.
- Consumer / Scientist does not want to know the underlying details.
  → His refrigerator also just works.


Domain Apps
Domain Apps
… …
Domain Apps
Domain Apps
eScience Middleware
+ Machine Learning + Reasoning + Scheduling + …
Service Plane
Chromium CGLX
SAGE MTP
OCCI JSDL
GIR UR
SNMP OpenFlow NSI
PerfSonar ICMP
Cassandra iRODs OGSA-DAIS
Hadoop Storm
WSRF SensorML
WebServ INSPIRE
GRID/Cloud Computing

Data intensive applications on programmable infrastructure: Intelligent workflow resource planning on the Network Service Interface (NSI)

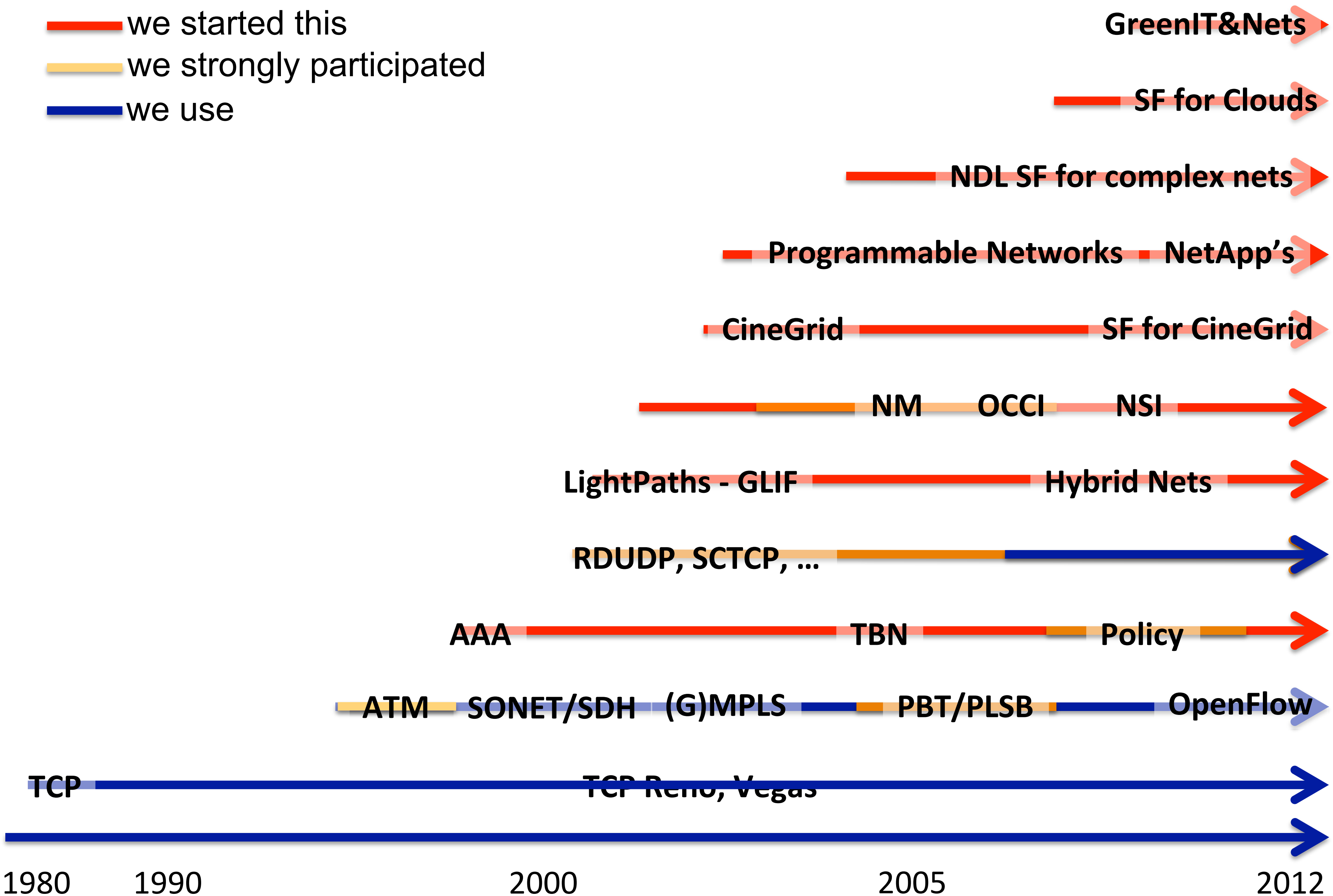
System and Network Engineering





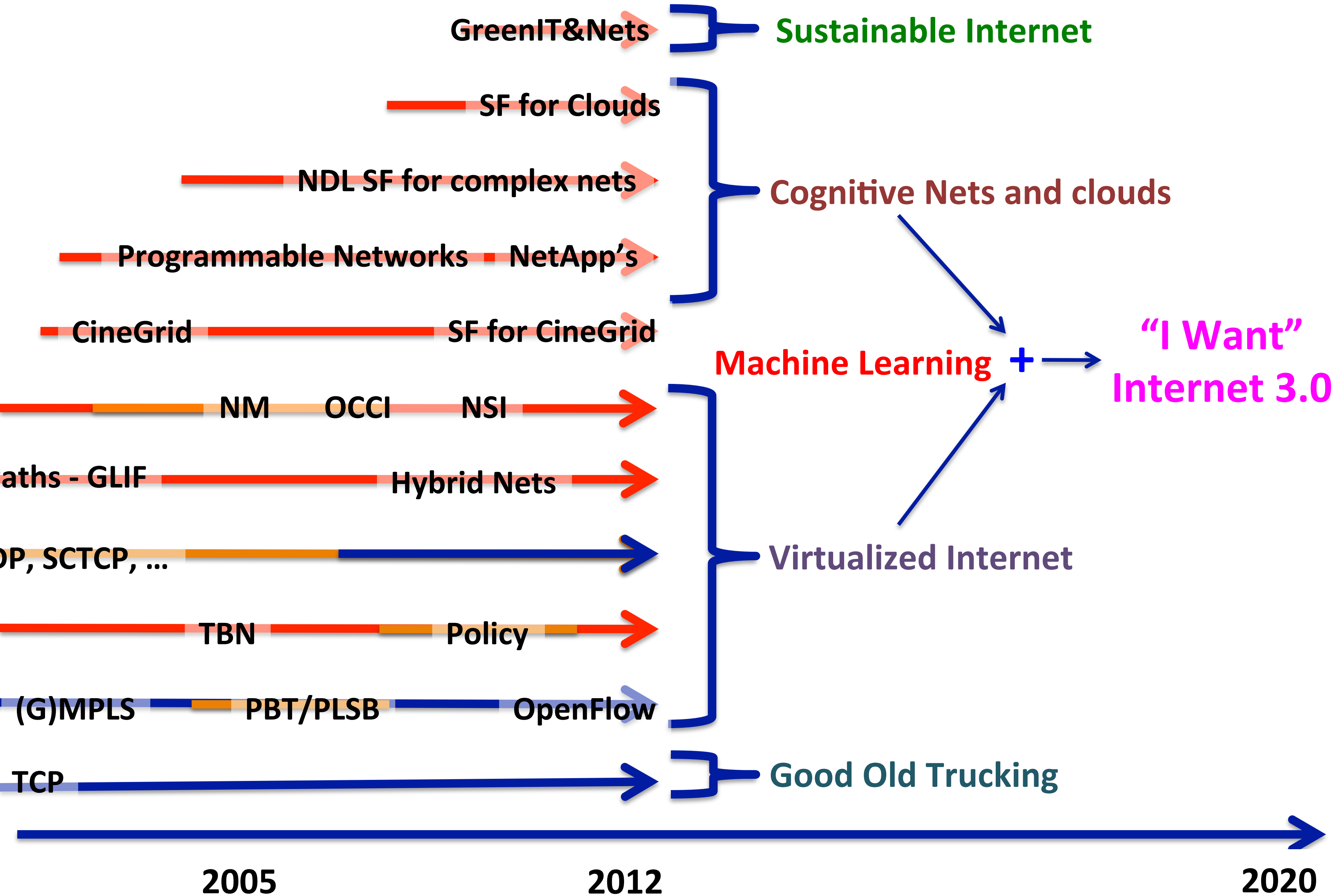
1. Zhao, Z., et al., (2012). Planning data intensive workflows on inter-domain resources using the Network Service Interface (NSI), WORKS in SC12
2. Zhao, Z., et al., (2011). An agent based network resource planner for workflow applications. International journal of Multiagent and Grid Systems, 7(6),

TimeLine



we started this
we strongly participated
we use
GreenIT&Nets
SF for Clouds
NDL SF for complex nets
Programmable Networks
NetApp's
CineGrid
SF for CineGrid
NM
OCCI
NSI
LightPaths - GLIF
Hybrid Nets
RDUDP, SCTCP, …
AAA
TBN
Policy
ATM
SONET/SDH
(G)MPLS
PBT/PLSB
OpenFlow
TCP
TCP Reno, Vegas
1980
1990
2000
2005
2012

TimeLine


GreenIT&Nets
Sustainable Internet
SF for Clouds
NDL SF for complex nets
Cognitive Nets and clouds
Programmable Networks
NetApp's
CineGrid
SF for CineGrid
Machine Learning
+
"I Want"
Internet 3.0
NM
OCCI
NSI
aths - GLIF
Hybrid Nets
P, SCTCP, ...
Virtualized Internet
TBN
Policy
(G)MPLS
PBT/PLSB
OpenFlow
TCP
Good Old Trucking
2005
2012
2020

TimeLine

Sustainable Internet

Cognitive Nets and clouds

lachine Learning + → “I Want” Internet 3.0

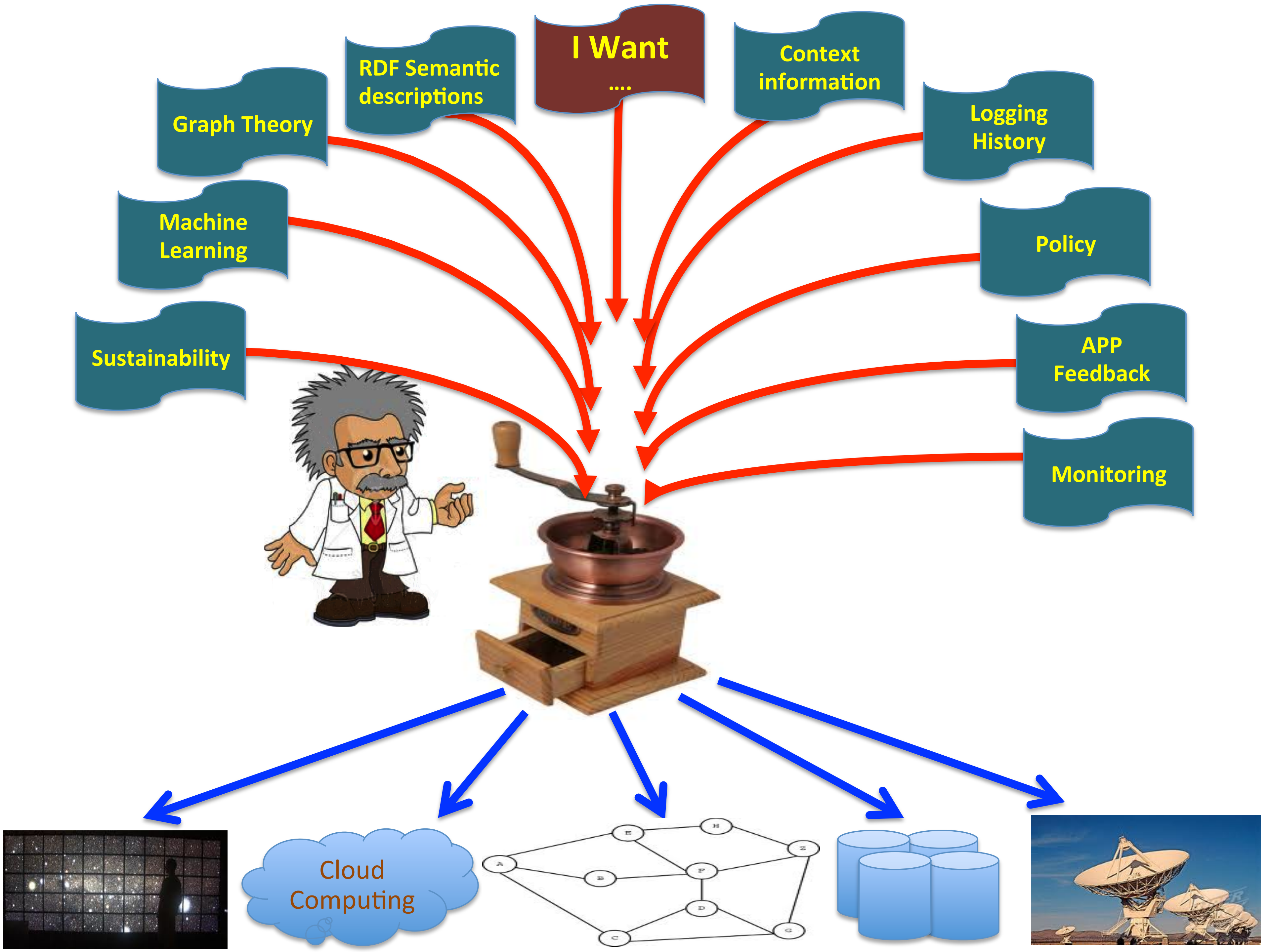
Virtualized Internet

Good Old Trucking

I retire

2020

2040

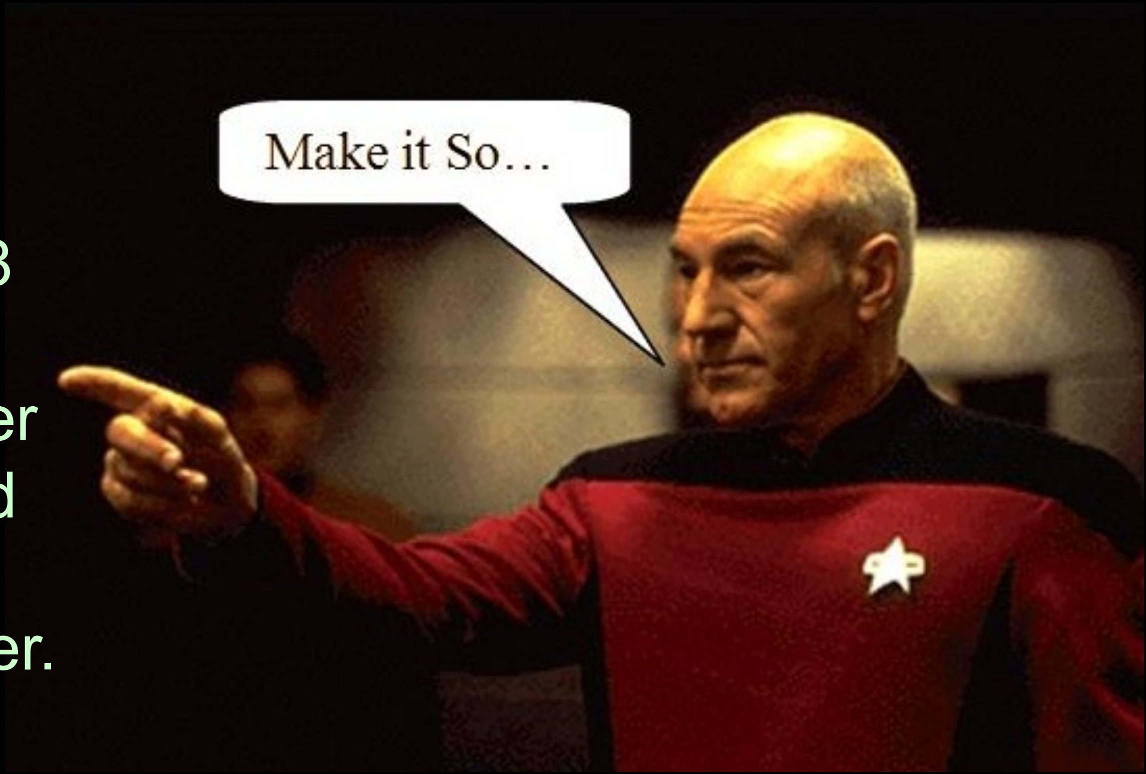


Graph Theory
RDF Semantic descriptions
I Want ….
Context information
Logging History
Machine Learning
Policy
Sustainability
APP Feedback
Monitoring
Cloud Computing
A
B
C
D
E
F
G
H
Z

Conclusion

I want a MiS system!

Catchphrase first used in "Encounter At Farpoint" (28 September 1987) by Gene Roddenberry, and thereafter used in many episodes and films, instructing a crew member to execute an order.



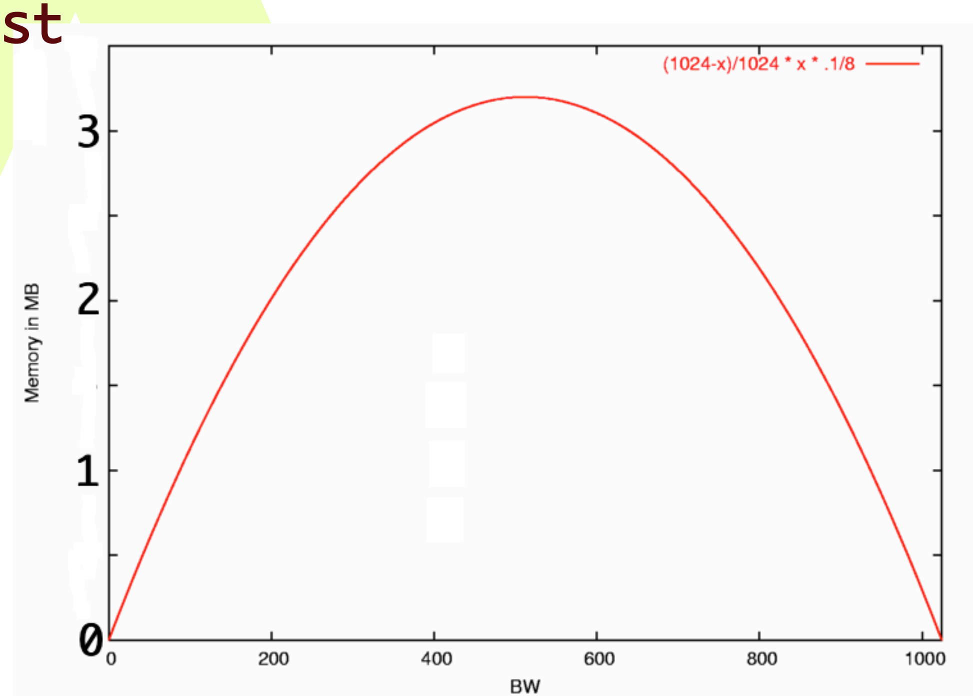
Layer - 2 requirements from 3/4

TCP is bursty due to sliding window protocol and slow start algorithm.

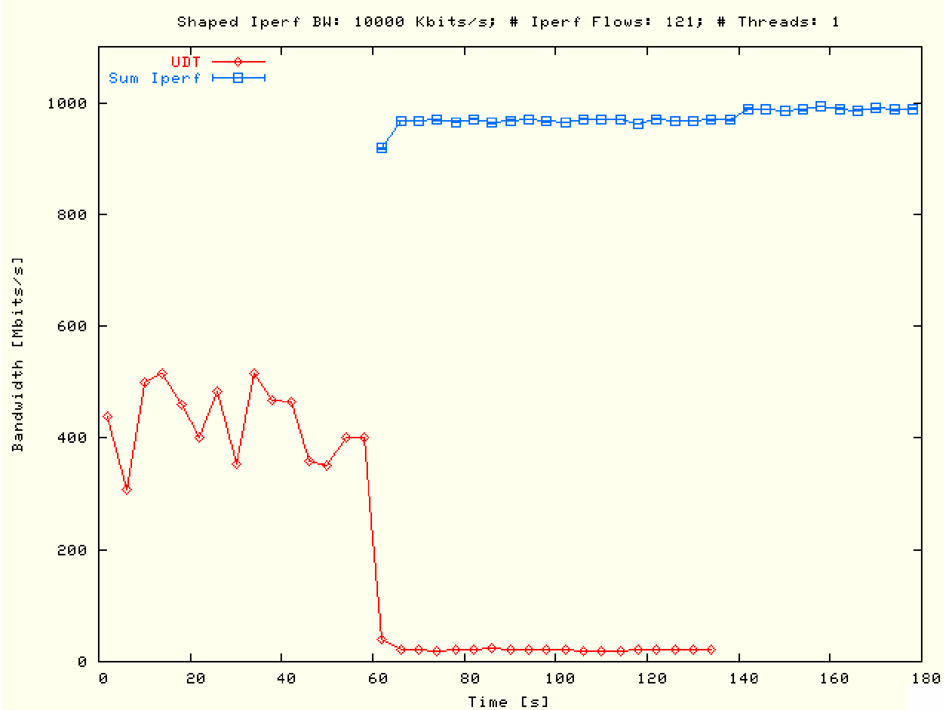
```
Window = BandWidth * RTT    &   BW == slow

                          fast - slow
Memory-at-bottleneck = ------------ * slow * RTT
                              fast
```

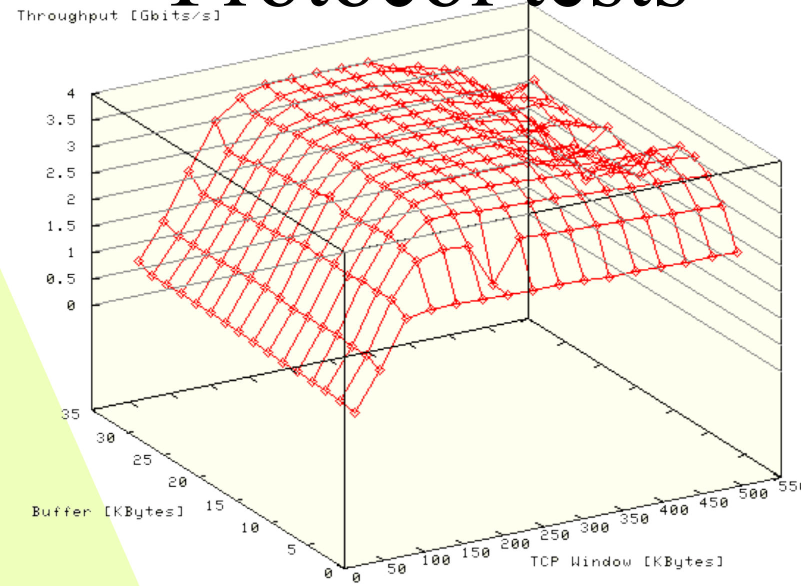
So pick from menu:

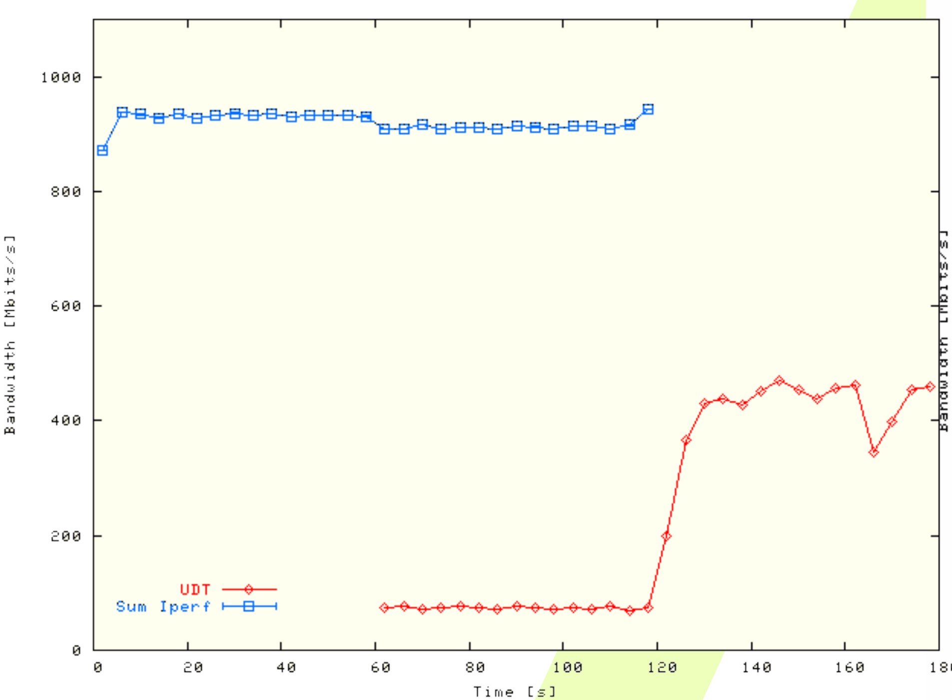
- *Flow control*
- *Traffic Shaping*
- *RED (Random Early Discard)*
- *Self clocking in TCP*
- *Deep memory*

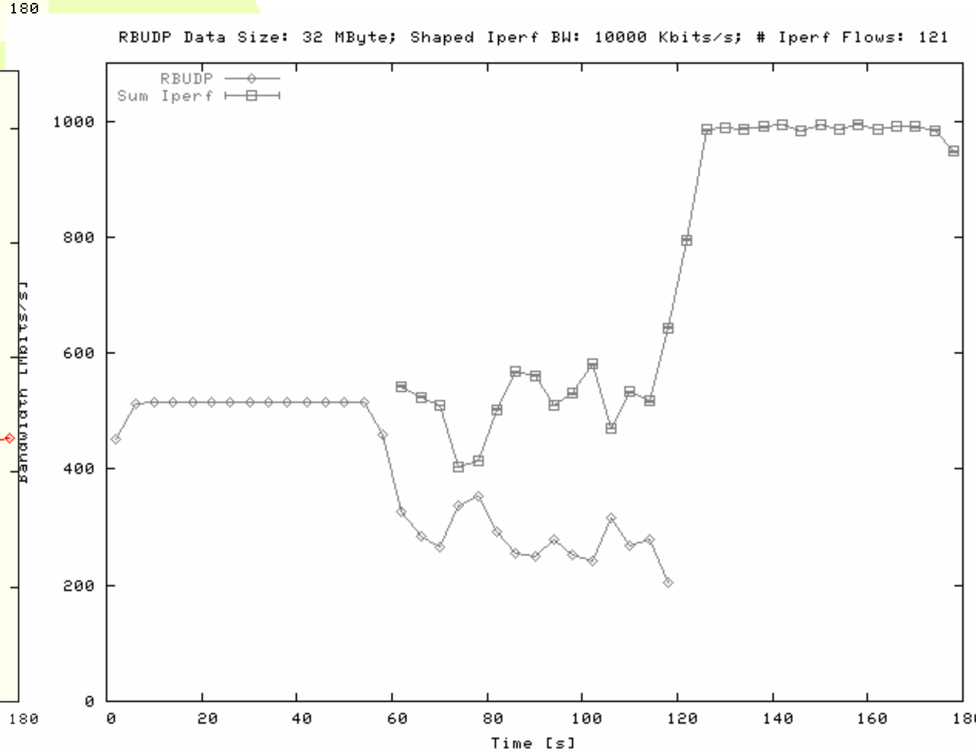


Protocol tests



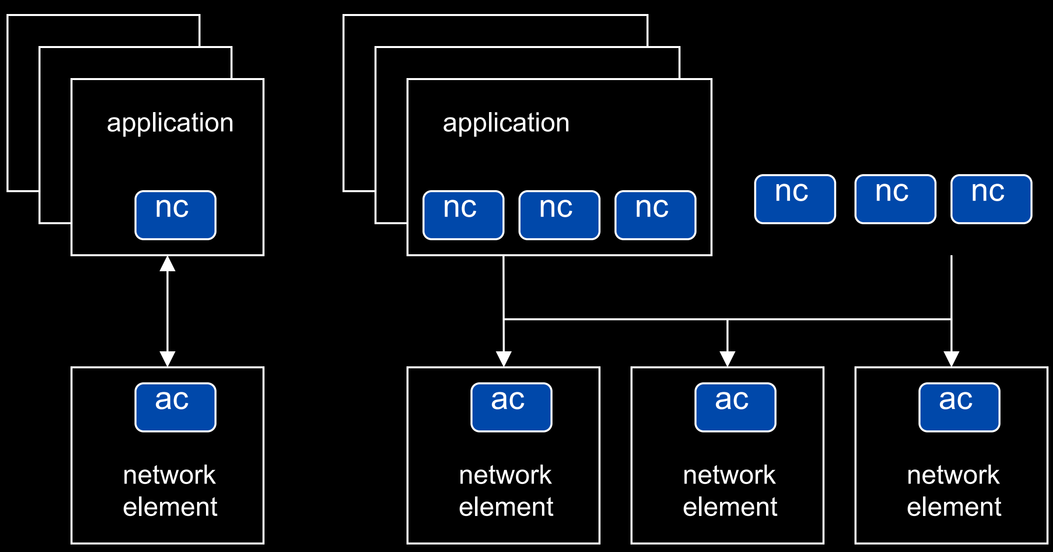


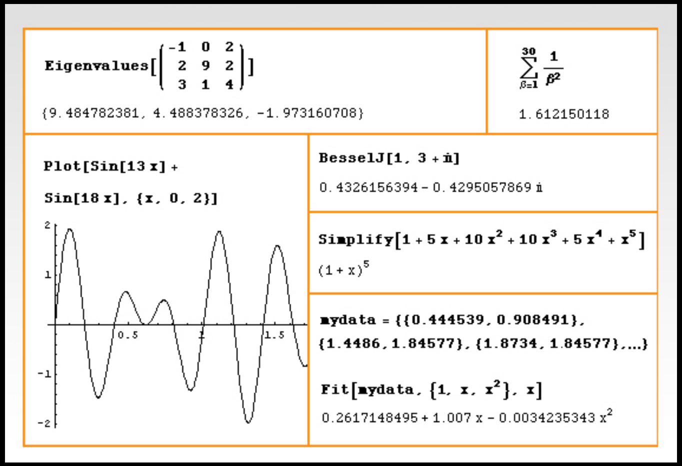




User Programmable Virtualized Networks allows the results of decades of computer science to handle the complexities of application specific networking.

- The network is virtualized as a collection of resources
- UPVNs enable network resources to be programmed as part of the application
- Mathematica, a powerful mathematical software system, can interact with real networks using UPVNs





StarPlane

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

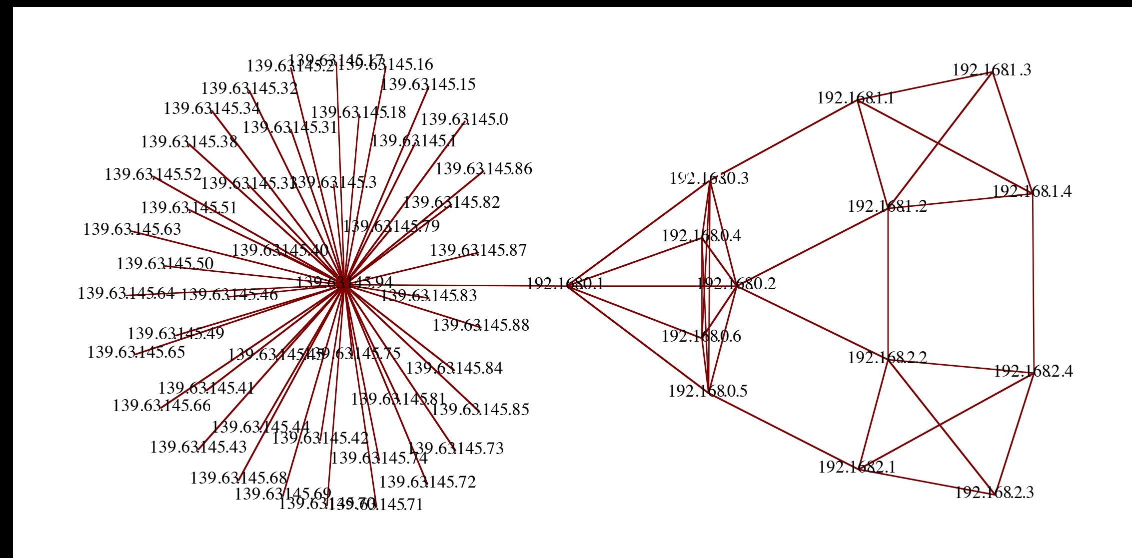
Initialization and BFS discovery of NEs

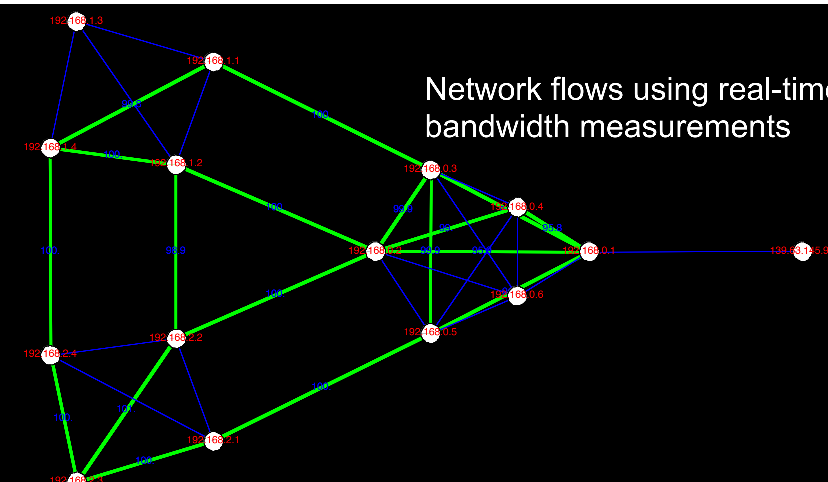
```
Needs["WebServices`"]
<<DiscreteMath`Combinatorica`
<<DiscreteMath`GraphPlot`
InitNetworkTopologyService["edge.ict.tno.nl"]

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

```
Internal links: {192.168.2.3}
```

```
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.

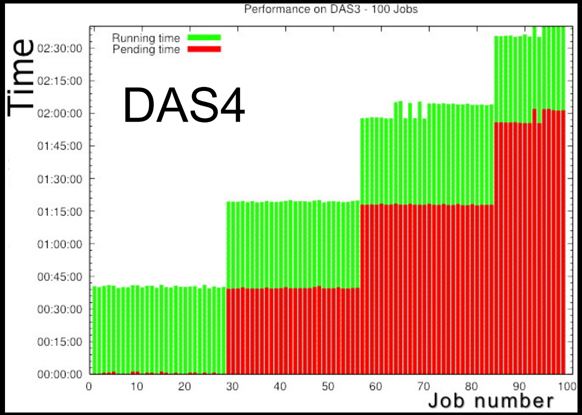
StarPlane

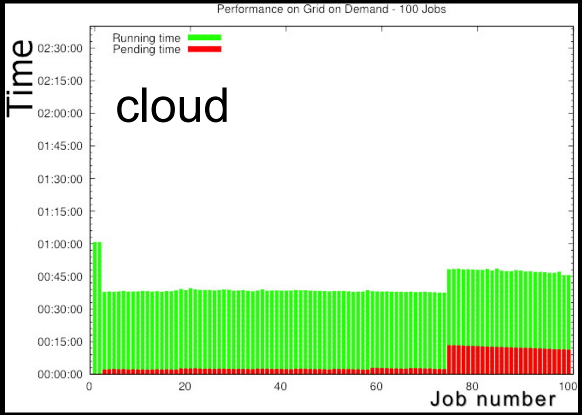
Demonstration of *optimizing the computing problem* ("Clouds")

If computing is 'infinite' and movable, then workflows and applications can ***program*** the network.

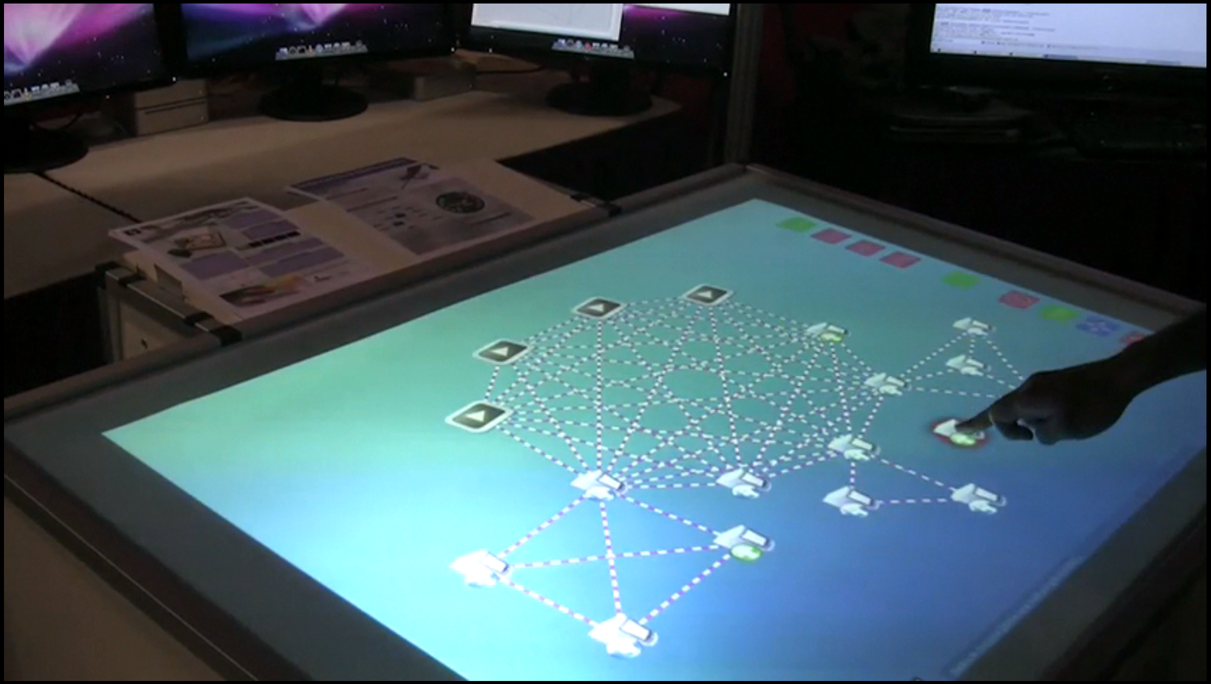
You can introduce new metrics when creating and optimizing these infrastructures (*e.g* power consumption)

Grid-on-demand





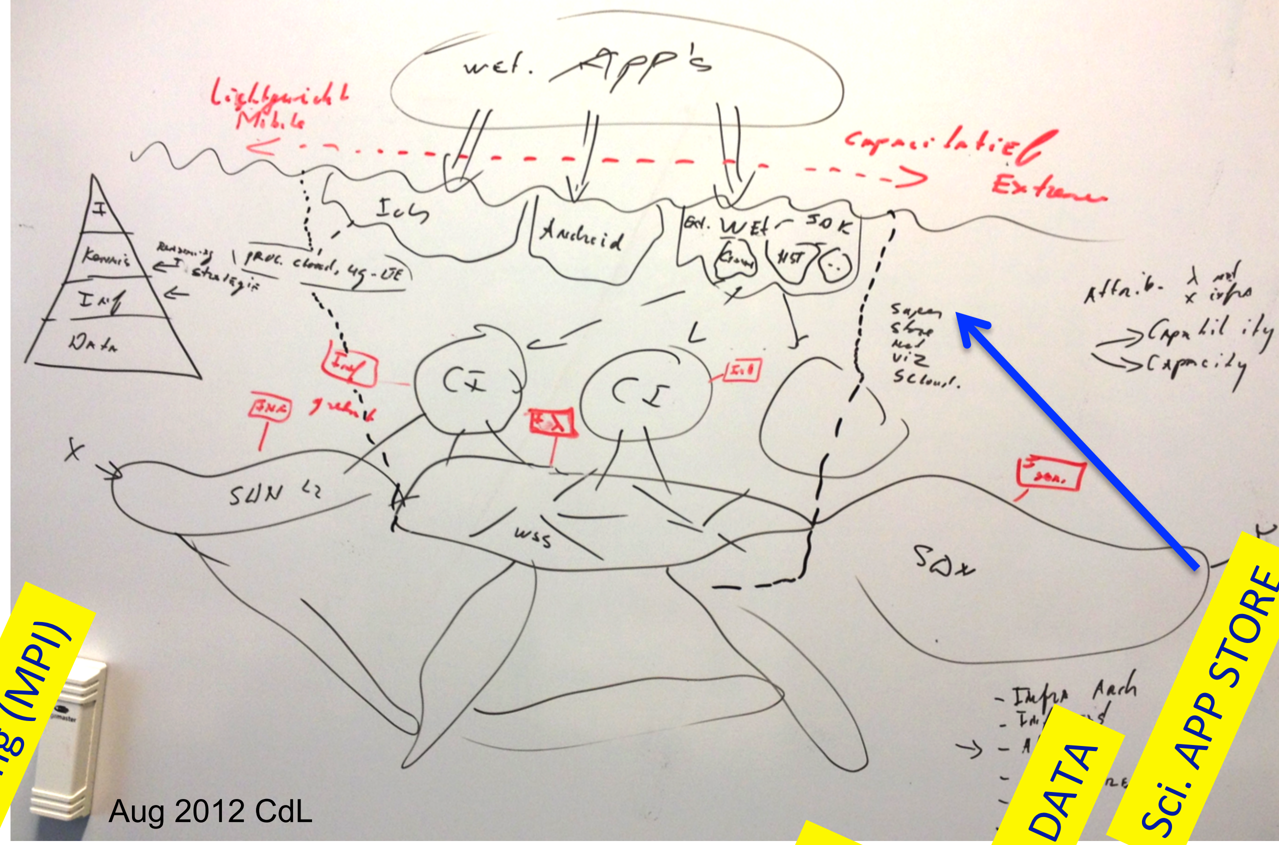
User programmable networks

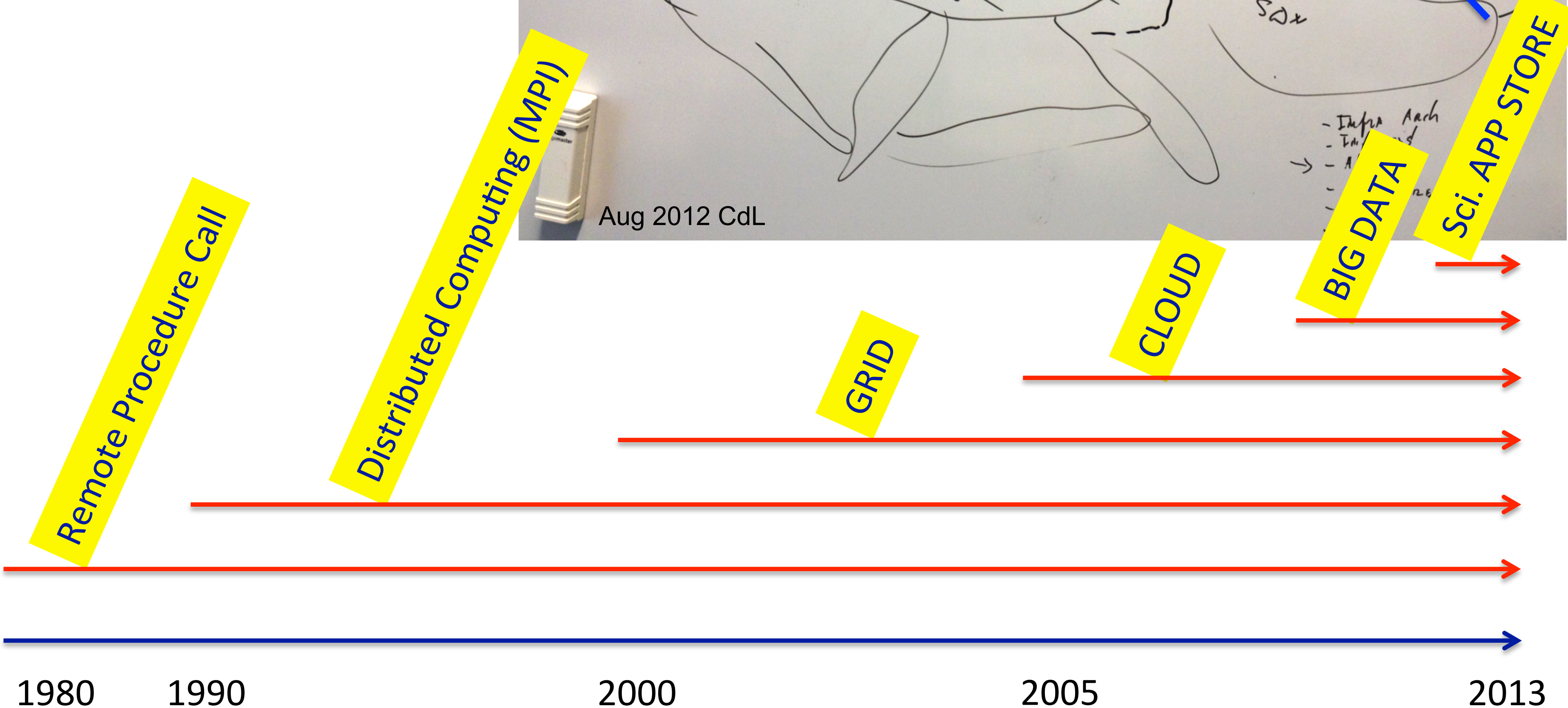


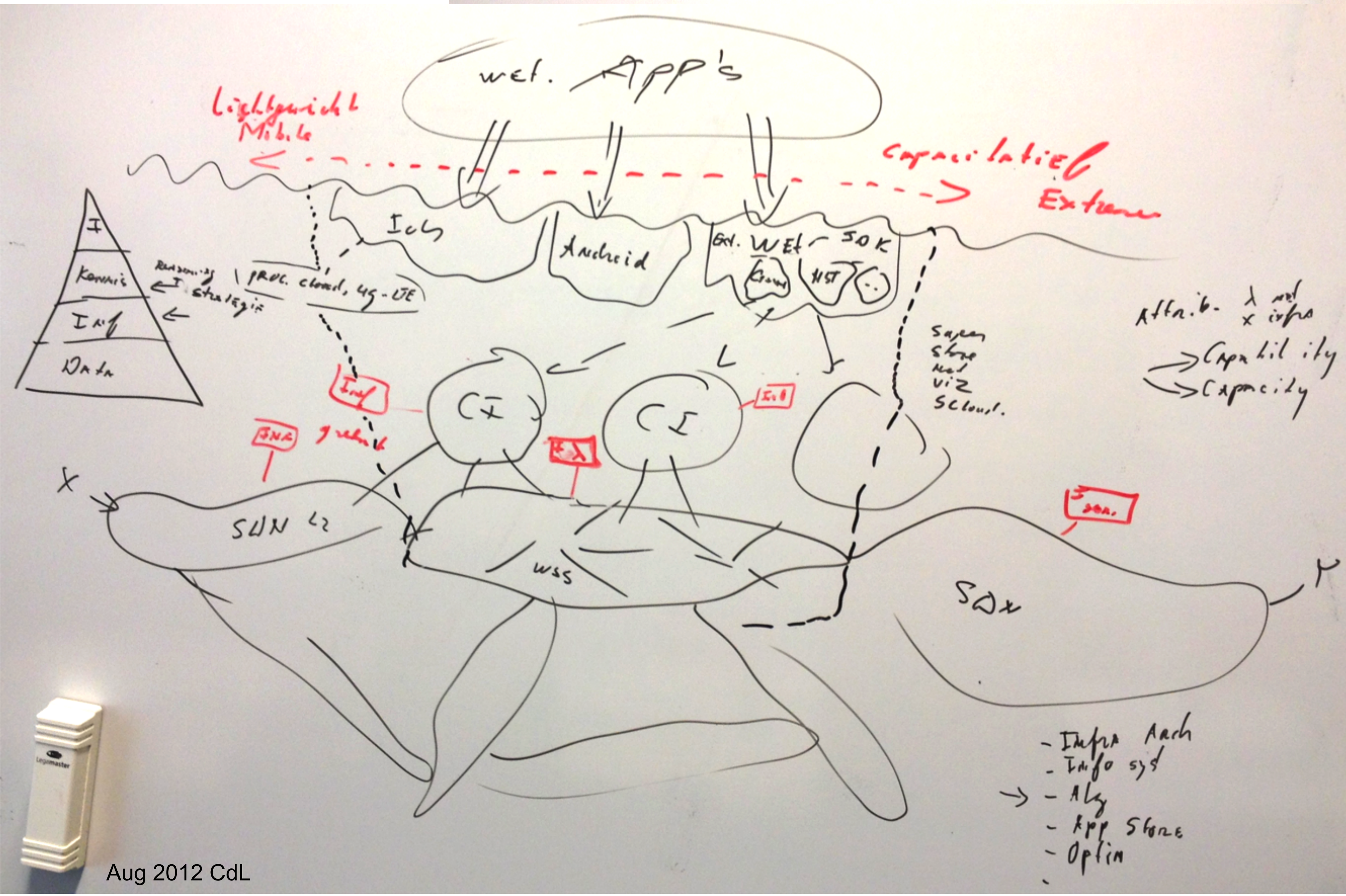
R.Strijkers, W.Toorop, A.van Hoof, P.Grosso, A.Belloum, D.Vasuining, C. de Laat, R. Meijer, "AMOS: Using the Cloud for On-Demand Execution of e-Science Applications", In: Proc. eScience2010 conf, Dec. 2010.

Y. Demchenko, C.Ngo, M.Makkes, R.Strijkers, C. de Laat, "Defining Inter-Cloud Architecture for Interoperability and Integration.", 3th intl conf on Cloud Computing, GRIDs, and Virtualization (CLOUDCOM 2012), July 22-27, 2012, Nice, France. BEST PAPER AWARD

TimeLine


Aug 2012 CdL

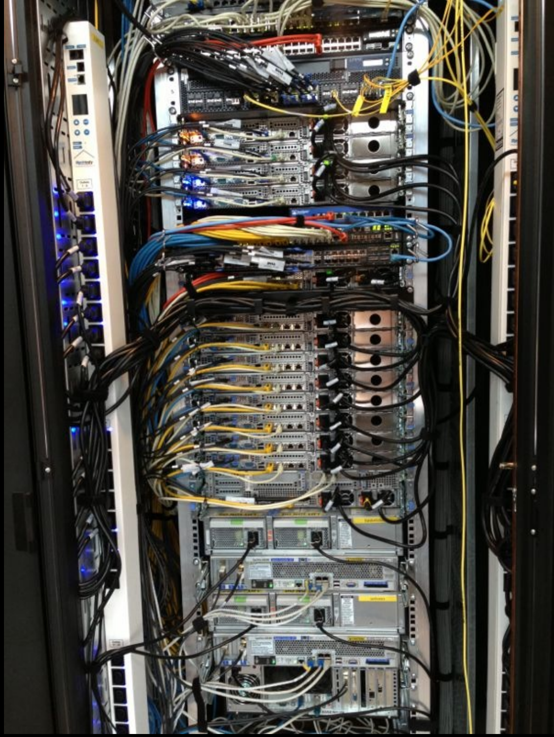

Remote Procedure Call
Distributed Computing (MPI)
GRID
CLOUD
BIG DATA
Sci. APP STORE
1980
1990
2000
2005
2013

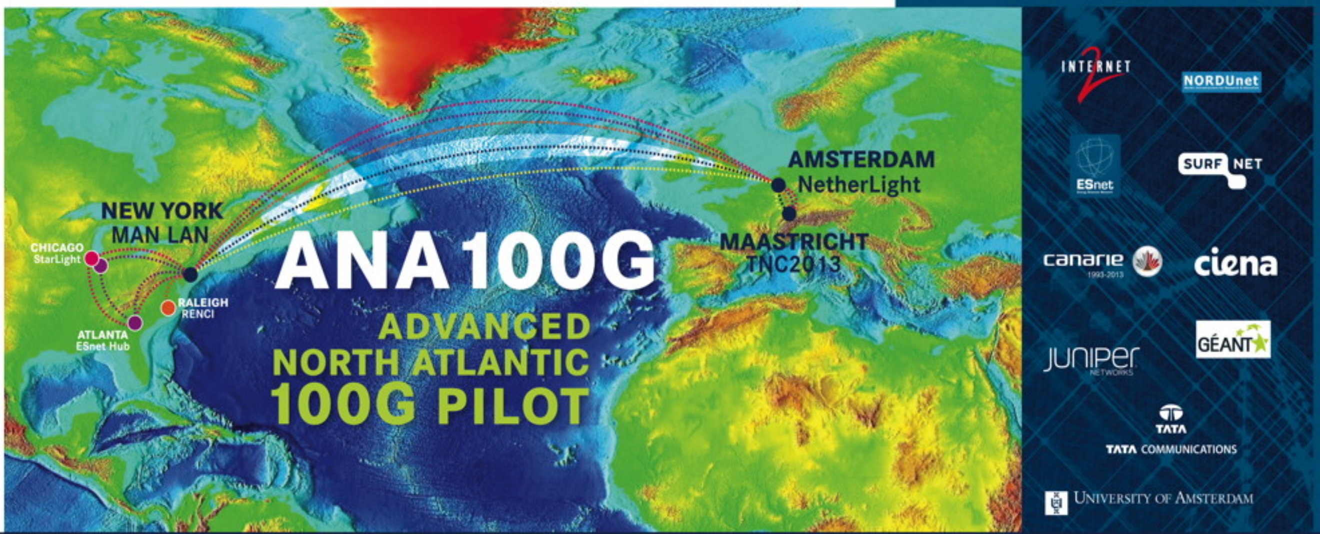


1980 1990 2000 2005 2013

ExoGeni @ UvA

Installed and up June 3th 2013





TNC2013 DEMOS JUNE, 2013

DEMO	TITLE	OWNER	AFFILIATION	E-MAIL	A-SIDE	Z-SIDE	PORTS(S) MAN LAN	PORTS(S) TNC2013	DETAILS
1	Big data transfers with multipathing, OpenFlow and MPTCP	Ronald van der Pol	SURFnet	ronald.vanderpol@surfnet.nl	TNC/MECC, Maastricht NL	Chicago, IL	Existing 100G link between internet2 and ESnet	2x40GE (Juniper)+ 2x10GE (OME6500)	In this demonstration we show how multipathing, OpenFlow and Multipath TCP (MPTCP) can help in large file transfers between data centres (Maastricht and Chicago). An OpenFlow application provisions multiple paths between the servers and MPTCP will be used on the servers to simultaneously send traffic across all those paths. This demo uses 2x40G on the transatlantic 100G link. ESnet provides 2x40G between MAN LAN and StarLight, ACE and USLHCnet provide additional 10GEs.
2	Visualize 100G traffic	Inder Monga	ESnet	imonga@es.net					Using an SNMP feed from the Juniper switch at TNC2013,and/or Brocade AL2S node in MANLAN, this demo would visualize the total traffic on the link, of all demos aggregated. The network diagram will show the transatlantic topology and some of the demo topologies.
3	How many modern servers can fill a 100Gbps Transatlantic Circuit?	Inder Monga	ESnet	imonga@es.net	Chicago, Ill	TNC showfloor	1x 100GE	8x 10GE	In this demonstration, we show that with the proper tuning and tool, only 2 hosts on each continent can generate almost 80Gbps of traffic. Each server has 4 10G NICS connected to a 40G virtual circuit, and has iperf3 running to generate traffic. ESnet's new "iperf3" throughput measurement tool, still in 'beta', combines the best features from other tools such as iperf, nuttcp, and netperf. See: https://my.es.net/demos/tnc2013/
4	First European ExoGENI at Work	Jeroen van der Ham	UvA	vdham@uva.nl	RENCI, NC	UvA, Amsterdam, NL	1x 10GE	1x 10GE	The ExoGENI racks at RENCI and UvA will be interconnected over a 10G pipe and be on continuously, showing GENI connectivity between Amsterdam and the rest of the GENI nodes in the USA.
5	Up and down North Atlantic @ 100G	Michael Enrico	DANTE	michael.enrico@dante.net	TNC showfloor	TNC showfloor	1x 100GE	1x 100GE	The DANTE 100GE test set will be placed at the TNC2013 showfloor and connected to the Juniper at 100G. When this demo is running a loop @ MAN LAN's Brocade switch will ensure that the traffic sent to MAN LAN returns to the showfloor. On display is the throughput and RTT (to show the traffic traveled the Atlantic twice)

Connected via the new 100 Gb/s transatlantic

ExoGeni @ UvA

- Part of UvA's OpenLab → Open for everyone!
- Installed and up June 3th 2013
- Connected via the new 100 Gb/s transatlantic
- To study programmability on all layers
- To study computing to data vs data to computing
- To study GreenSonar & objective based networking
- Study multi service exchange & DMZ features
- To study Big Data processing algorithms on mixed latency
- PIRE project with Grossman and Alvares
- Give students access to try out their bright and stupid ideas!
- DAS4/5, CineGrid exchange node, pure photonic TUE

The constant factor in our field is Change!

The 50 years it took Physicists to find one particle, the Higgs, we came from:

“Fortran goto”, Unix, c, SmallTalk, DECnet, TCP/IP, c++, Internet, WWW, Semantic Web, Photonic networks, Google, grid, cloud, Data^3, App

to:

DDOS attacks destroying Banks and Bitcoins.

Conclusion:

Need for Safe, Smart, Resilient Sustainable Infrastructure.

Many thanks to RENCI, CIENA, SURFnet, DELL,

Ralph, Jeroen, Daniel, Erik-Jan, Joe.

Questions

- This trip is supported by:
  - COMMIT WP 20.1

COMMIT/