Globally Distributed Secure Data Exchange Fabrics

Cees de Laat

Systems and Networking Laboratory
University of Amsterdam

Contributions from:

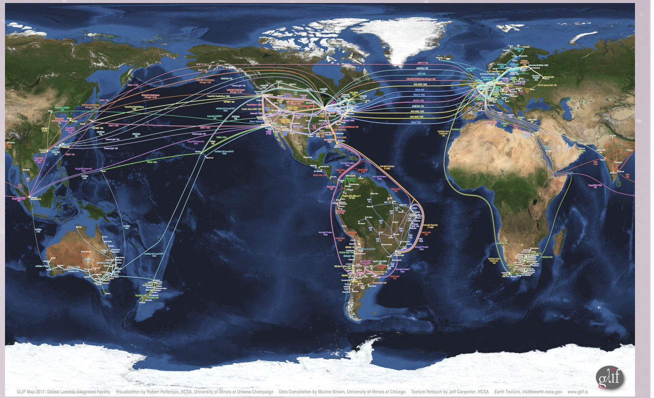
Leon Gommans, Paola Grosso, Wouter Los, Yuri Demchenko, Lydia Meijer, Tom van Engers, Sander Klous, Rodney Wilson, Marc Lyonais, Inder Monga, Reggie Cushing, Ameneh Deljoo, Sara Shakeri, Lu Zhang, Joseph Hill, Lukasz Makowski, Ralph Koning, Gleb Polevoy, Tim van Zalingen, and many others!

Mission

- The Systems and Networking Lab conducts research on leading-edge computer systems of all scales, ranging from global-scale systems and networks to embedded devices.
- Across these multiple scales our particular interest is on extra-functional properties of systems, such as performance, programmability, productivity, security, trust, sustainability and, last but not least, the societal impact of emerging systems-related technologies.

Broad spectrum of research

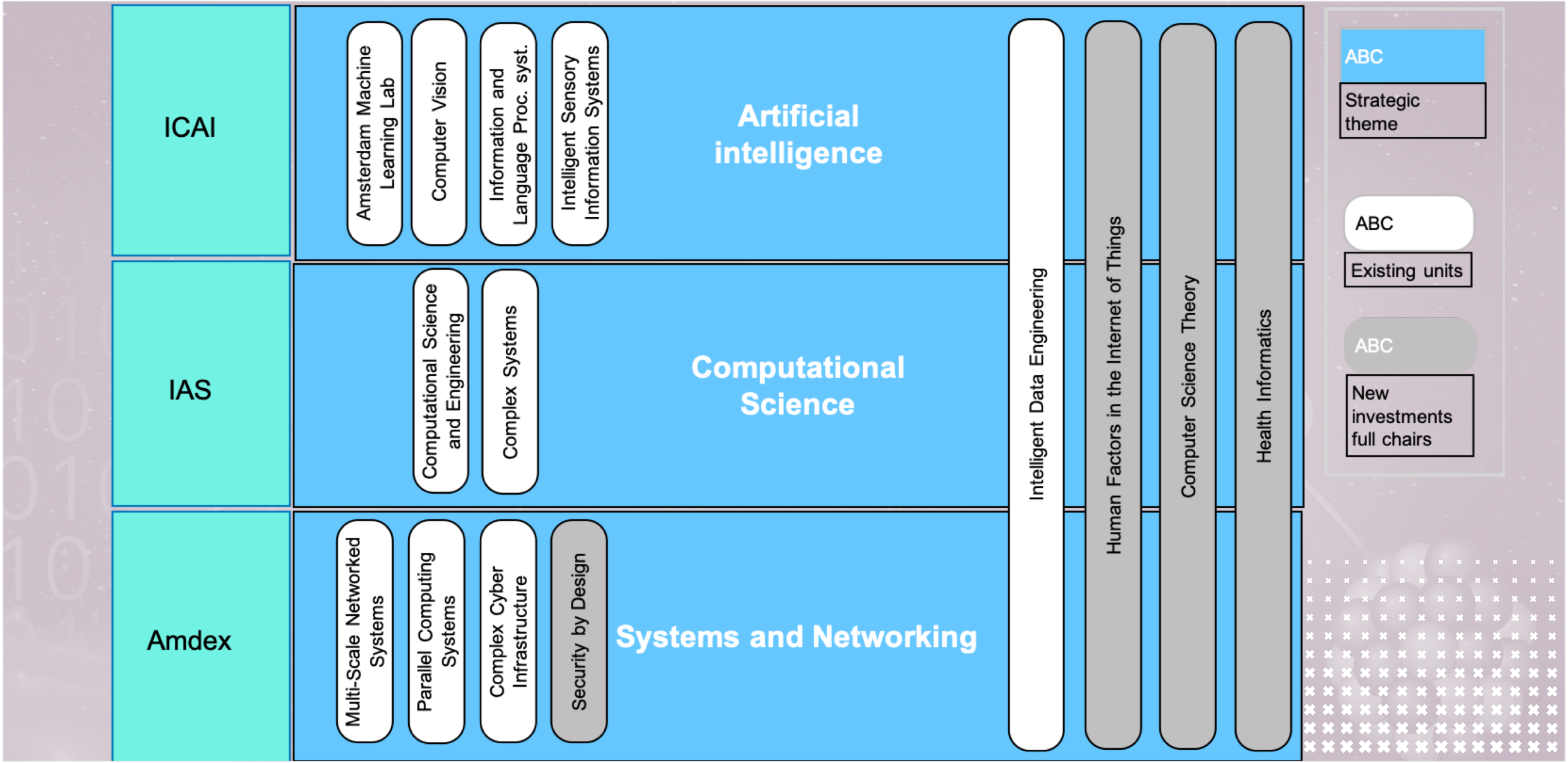
- Advanced networks / Internet architecture
- Network programmabillity / Overlays / Virtualisation
- Authorisation of Internet resources
- Quality of service for apps on Clouds
- Systems: Embedded / real time / parallel / design
- Performance & Compilers & ExaScale
- Safe Secure Data Sharing / Processing
- Data Sovereignty & Normative Agents & Trust ➔ AmDEX



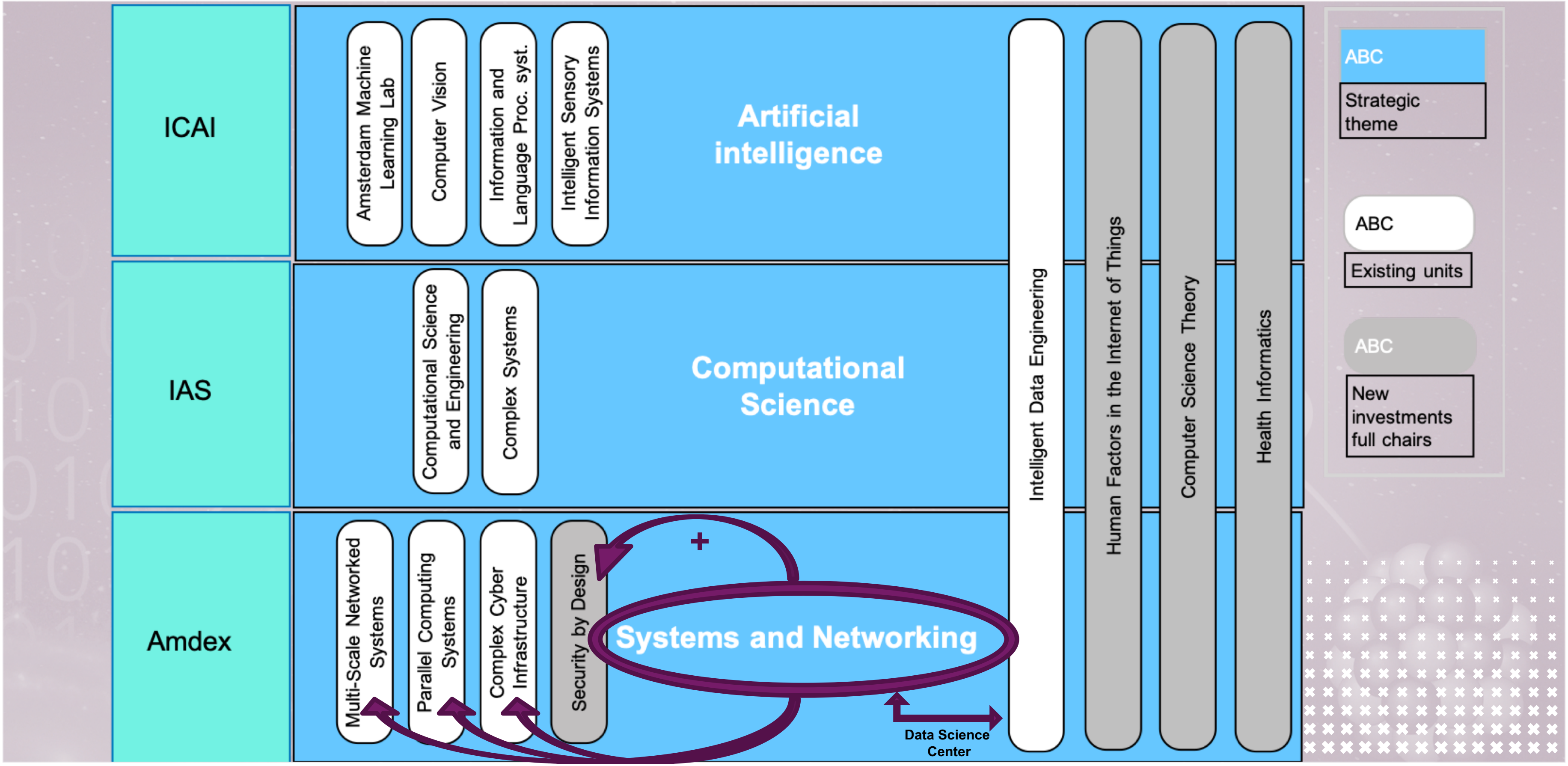
- **Well funded accross the themes ➔ ~ 10 Meuro in last 3 years, mostly from NWO / EU and smaller portion from SURF and Industry**

University of Amsterdam

Position in the Instituut



Position in the Instituut



Totals staffing in the SNE Lab

- Group leaders: prof.dr.ir. C. de Laat, dr. Paola Grosso, dr. Andy Pimentel
  - 1 full prof (CdL)
  - 4 associate professors (PG, AV, AP, CG)
  - 2+3 assistant professors (AB, ZZ, 3 open positions)
  - 3 part time professors (PA, PK, TvE)
  - 3 endowed professors (RvN, SK, LG)
  - 1 senior researcher (YD)
  - ~ 23 phd students, ~ 6 postdoc's
  - 2 tech support/scientific programmers
  - ~18 guests
  - Total about 65 people ~ 40 fte
- Yearly turnover ~ 3 MEuro research + similar education

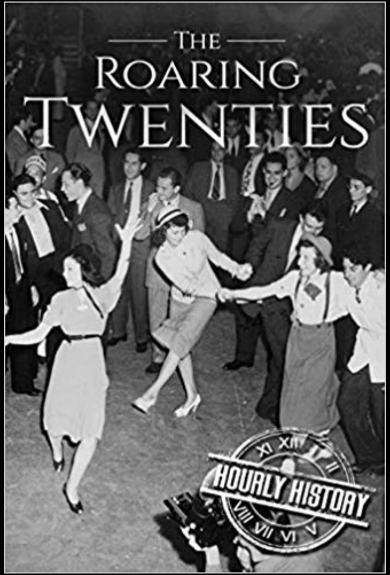
Research based Education

- The SNE laboratory is closely working together with the SNE Master (www.os3.nl), Software Engineering Master (SE) and the Computer Science Master (CS) programs to disseminate knowledge through education.

- **Both SNE and SE masters score in the top nationwide**

- **SNE master Excellent in accreditation and**

ICT to support the transformation of Science in the Roaring Twenties



From Wikipedia: The Roaring Twenties refers to the decade of the 1920s in Western society and Western culture. It was a period of economic prosperity with a distinctive cultural edge in the United States and Western Europe, particularly in major cities such as Berlin, Chicago, London, Los Angeles, New York City, Paris, and Sydney. In France, the decade was known as the "années folles" ('crazy years'), emphasizing the era's social, artistic and cultural dynamism. Jazz blossomed, the flapper redefined the modern look for British and American women, and Art Deco peaked….

This period saw the large-scale development and use of automobiles, telephones, movies, radio, and electrical appliances being installed in the lives of thousands of Westerners. Aviation soon became a business. Nations saw rapid industrial and economic growth, accelerated consumer demand, and introduced significantly new changes in lifestyle and culture. The media focused on celebrities, especially sports heroes and movie stars, as cities rooted for their home teams and filled the new palatial cinemas and gigantic sports stadiums. In most major democratic states, women won the right to vote. The right to vote made a huge impact on society.

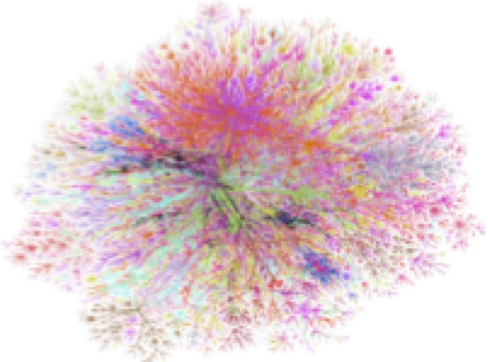
In most applications, utilization of Big Data often needs to be combined with Scalable Computing.

hadoop

NVIDIA TESLA

amazon web services

COMPUTING AT DIVERSE SCALES



"BIG" DATA

Enables dynamic data-driven applications

Computer-Aided Drug Discovery

Smart Cities

Disaster Resilience and Response

Smart Manufacturing

Personalized Precision Medicine

Smart Grid and Energy Management

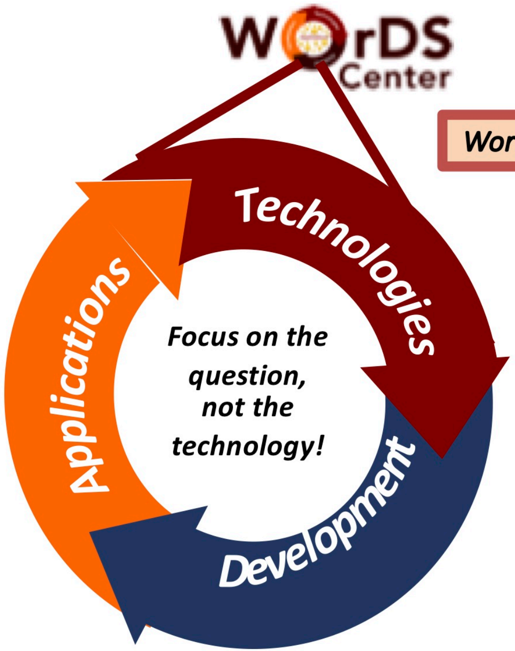
SDSC SAN DIEGO SUPERCOMPUTER CENTER

İlkay ALTINTAŞ, Ph.D.
ialtintas@ucsd.edu

UC San Diego

Workflows for Data Science Center of Excellence at SDSC

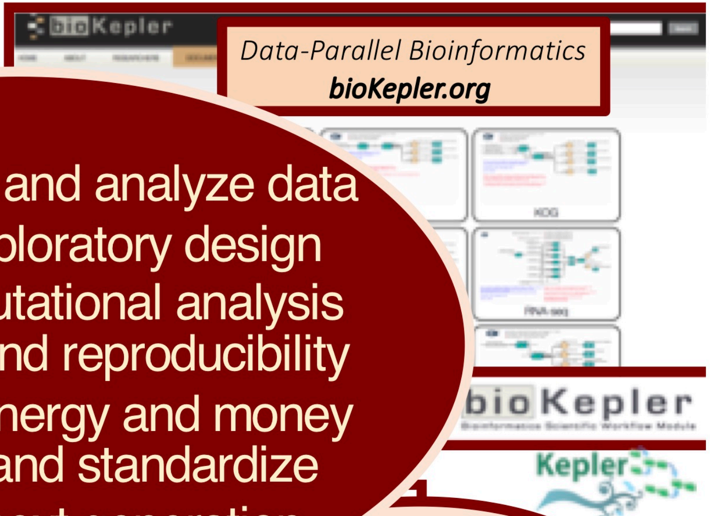
WorDS.sdsc.edu



Goal: Methodology and tool development to build automated and operational workflow-driven solution architectures on big data and HPC platforms.

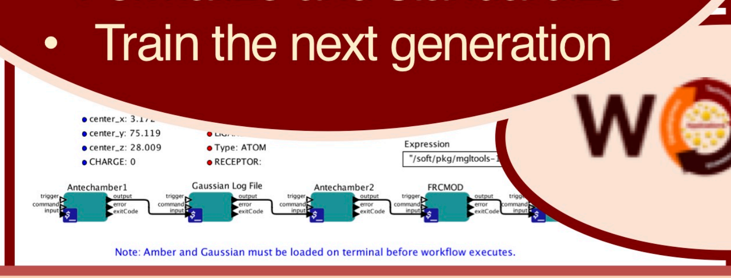
Real-Time Hazards Management
wifire.ucsd.edu

Data-Parallel Bioinformatics
bioKepler.org



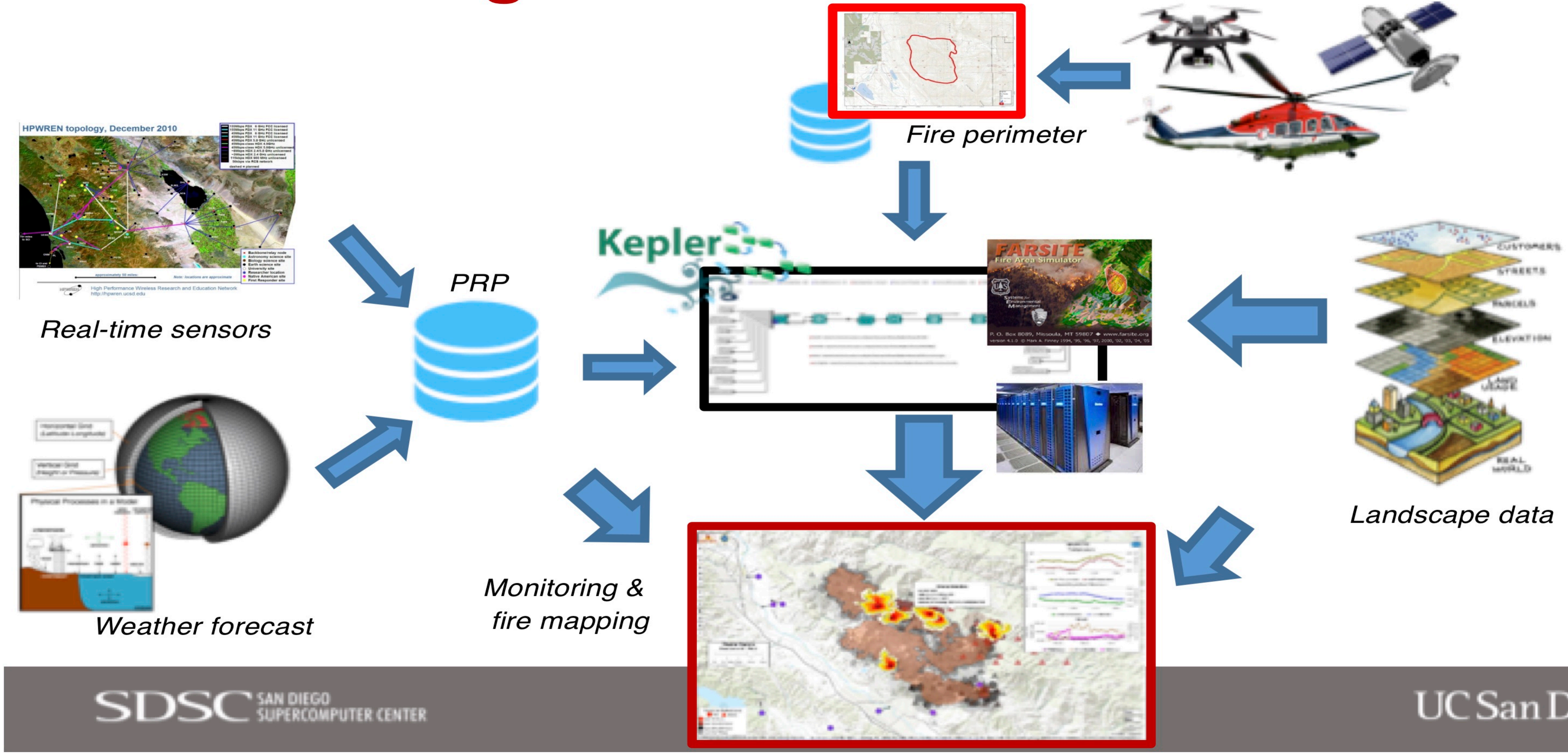
- Find, access and analyze data
- Support exploratory design
- Scale computational analysis
- Fuel reuse and reproducibility
- Save time, energy and money
- Formalize and standardize
- Train the next generation

kepler-project.org

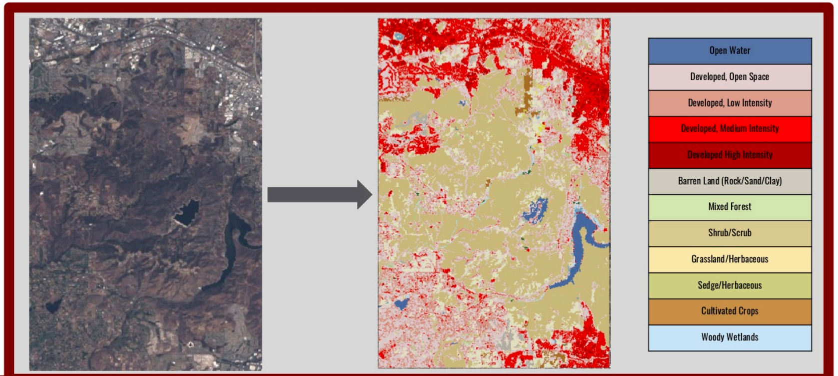


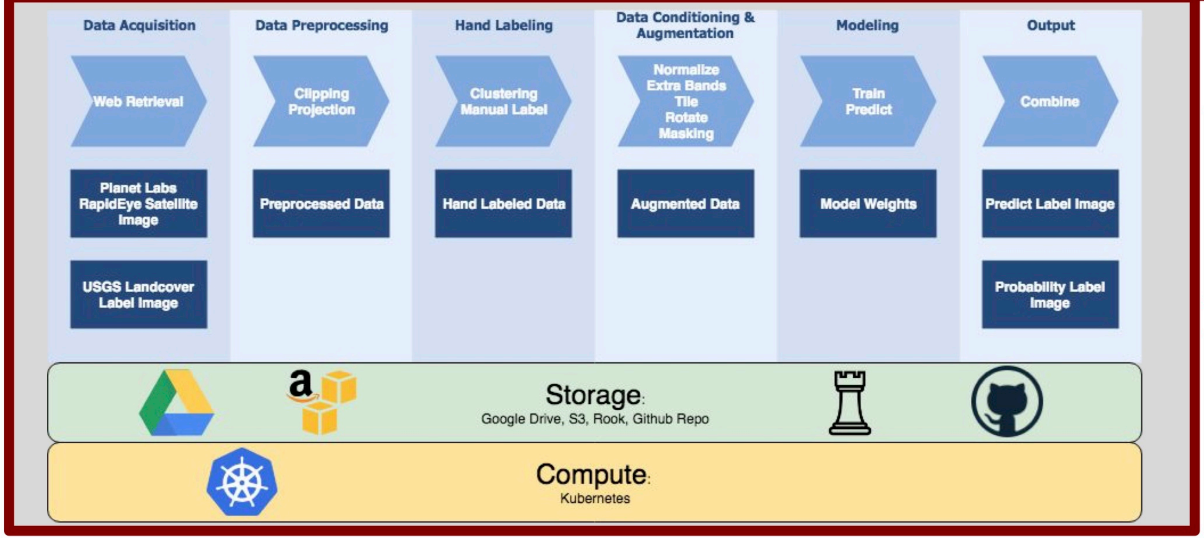
Scalable Automated Molecular Dynamics and Drug Discovery
nbcr.ucsd.edu

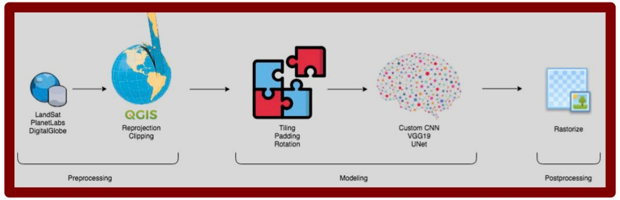
Fire Modeling Workflows in WIFIRE

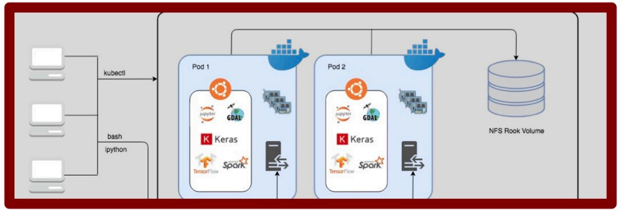


One Piece of the Puzzle: Vegetation Classification using Satellite Imagery





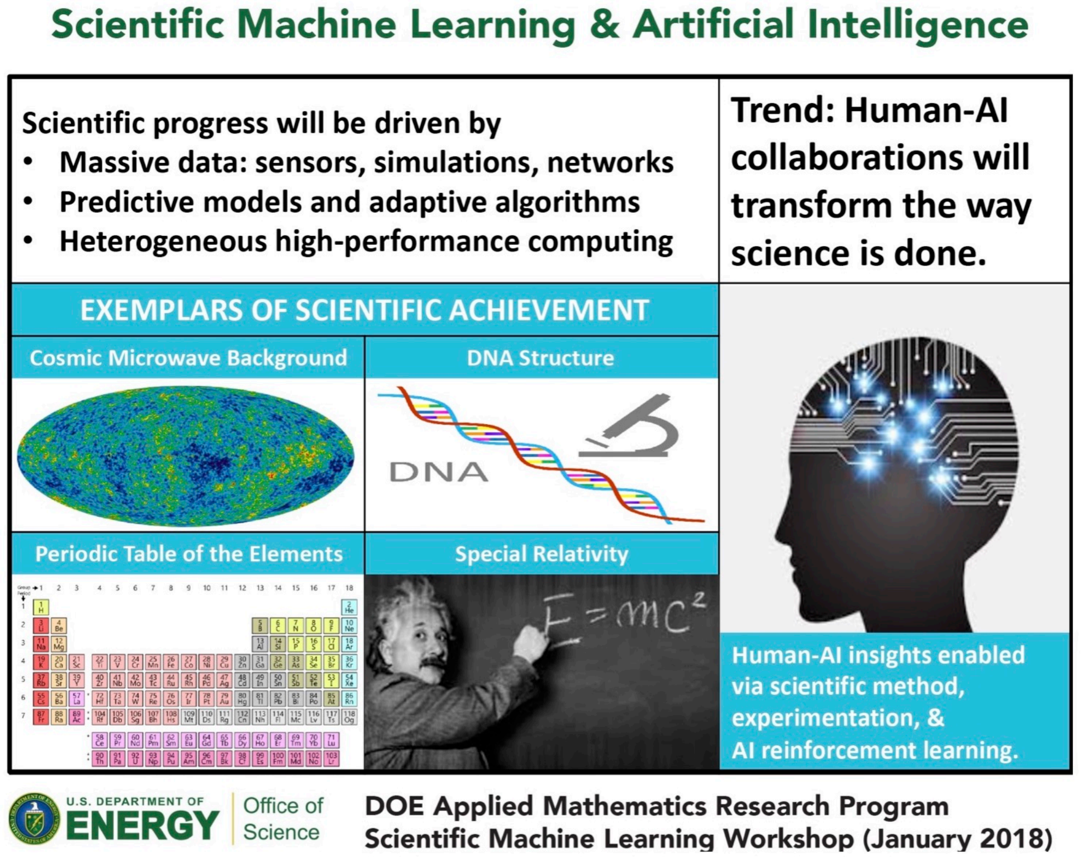




SDSC SAN DIEGO SUPERCOMPUTER CENTER

İlkay ALTINTAŞ, Ph.D.
ialtintas@ucsd.edu

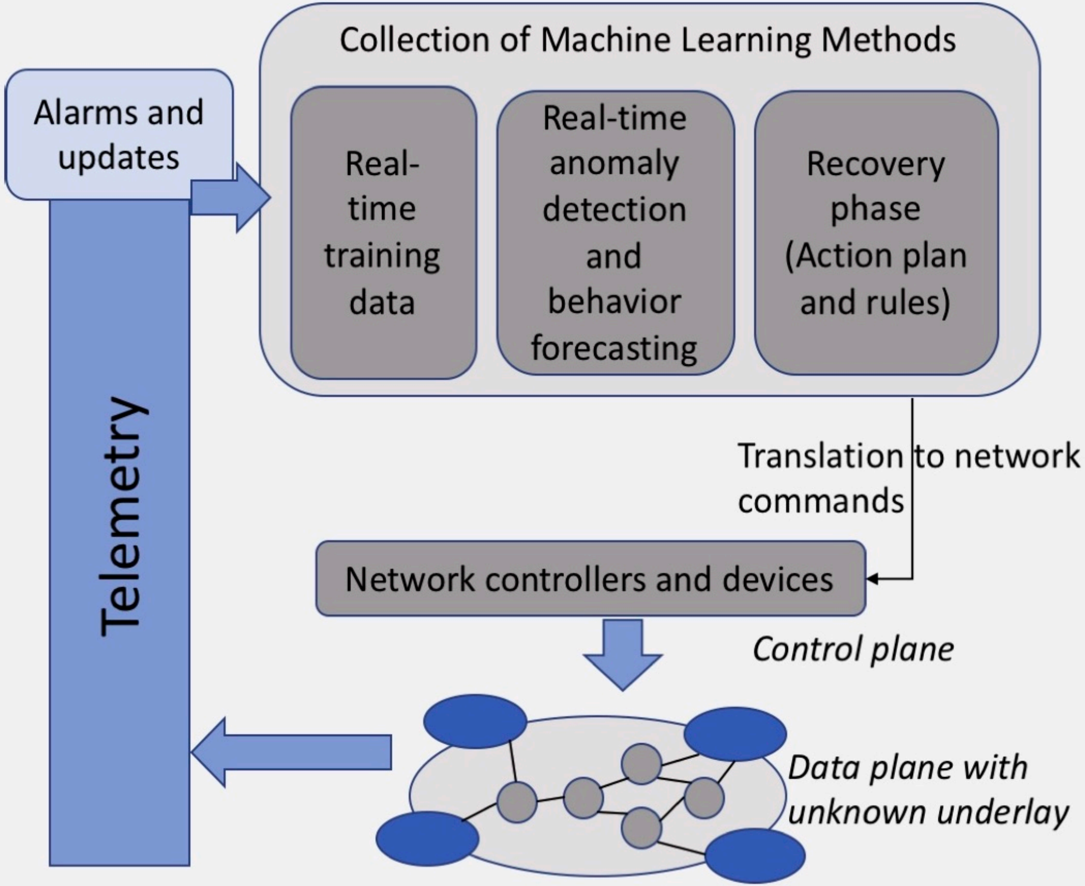
UC San Diego



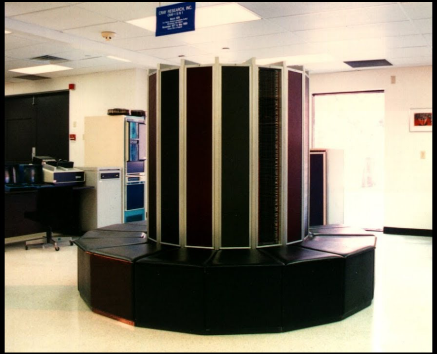
Workshop report:
https://www.osti.gov/biblio/1478744

Example 1: Optimizing Network Traffic with Machine Learning

Exascale and increasingly complex science applications are exponentially raising demands from underlying DOE networks, such as traffic management, operation scale, and reliability constraints. Networks are the backbone to complex science workflows, ensuring data are delivered securely and on time for important computations to happen. To optimize these distributed workflows, networks are required to understand end-to-end performance needs in advance and be faster, efficient, and more proactive, anticipating bottlenecks before they happen. However, to manage multiple network paths intelligently, various tasks, such as pre-computation and prediction, must be done in near real time. ML provides a collection of algorithms that can add autonomy and assist in decision making to sup-



Change in computing



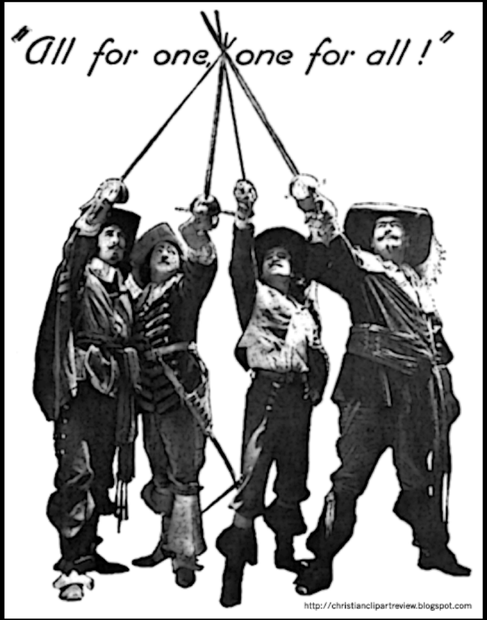
- Early days a few big Supercomputers
  - Mostly science domain
- Via grid to commercial cloud
  - AWS, Azure, Google Cloud, IBM, Salesforce
  - The big five: Apple, Alphabet, Microsoft, Facebook and Amazon
  - Computing has transformed into an utility

- Data => Information is the key

Main problem statement

- There is lots of data out there that is not shared (99%)
- FAIR is typically not fair ;-), but limited by policy and/or law
  - the A in FAIR is about access, trust is hard to implement across domains
- Organizations that normally compete have to bring data together to achieve a common goal/benefit!
- The shared data may be used for that goal but not for any other!
- Expected use is fine but unexpected use/mission creep…
- Data processed by alien algorithms in foreign data centers... Hmmm…
  - How to organize data processing alliances?
  - How to enforce policy using modern Cyber Infrastructure?
  - How to translate law policy from strategic via tactical to operational level?
  - What are the different fundamental data infrastructure models to consider?

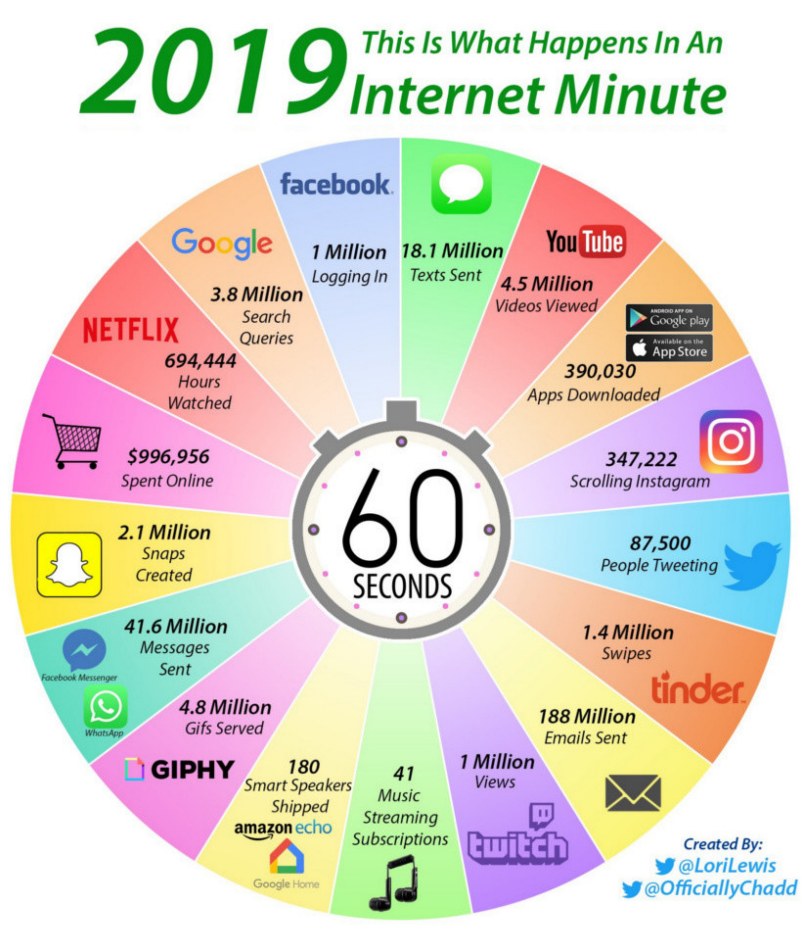
All for one and one for all



- All for one
  - Many infrastructures centered around compute and workflows
- One for all
  - Now we need to get a fluid data layer that frees data to be shared and used by (unforeseen) applications

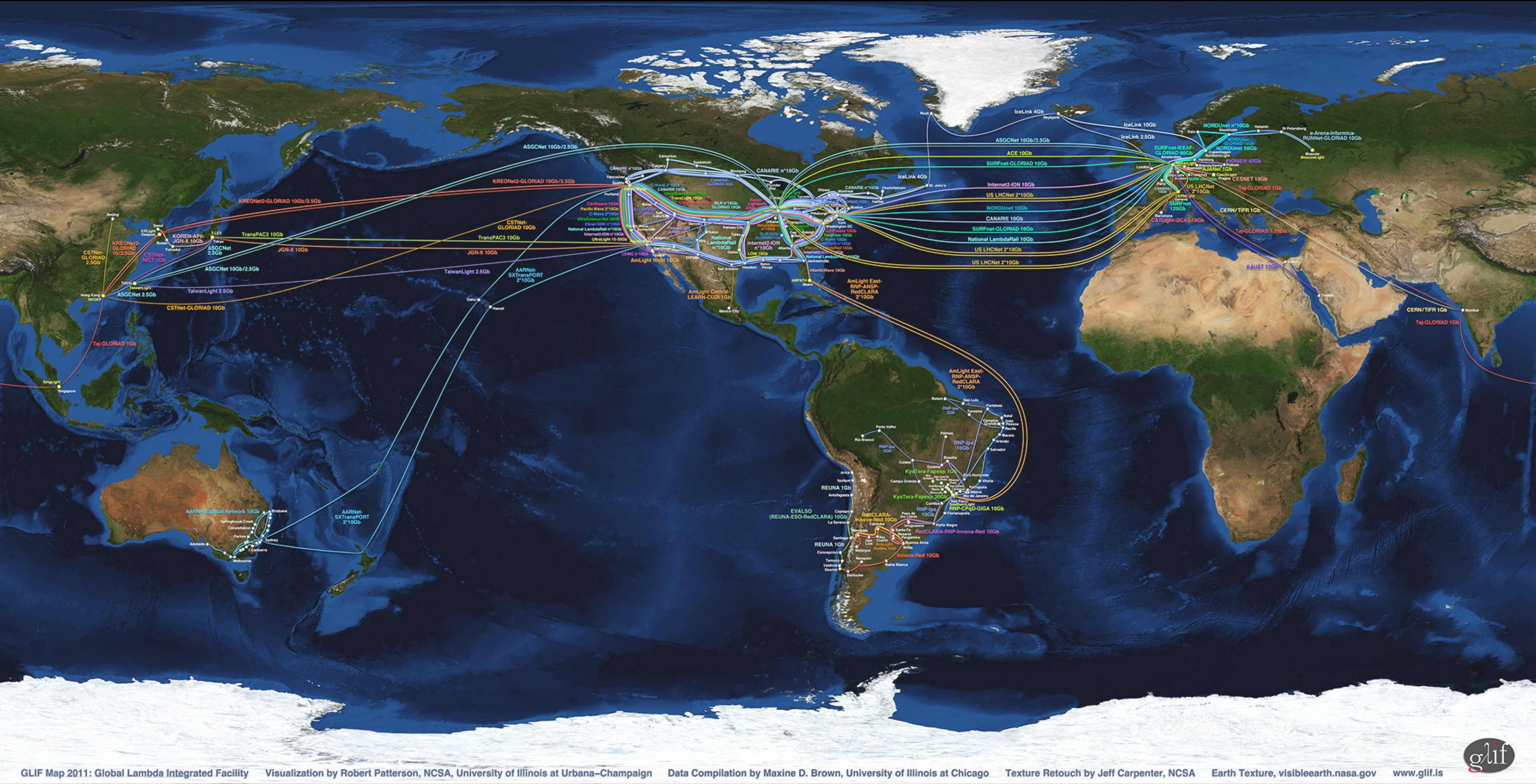
- Efforts as FAIR and ScienceDMZ / DTN fabrics pave the way to solve the data problem that is also encountered by industry.

Now, how do we get and use data?



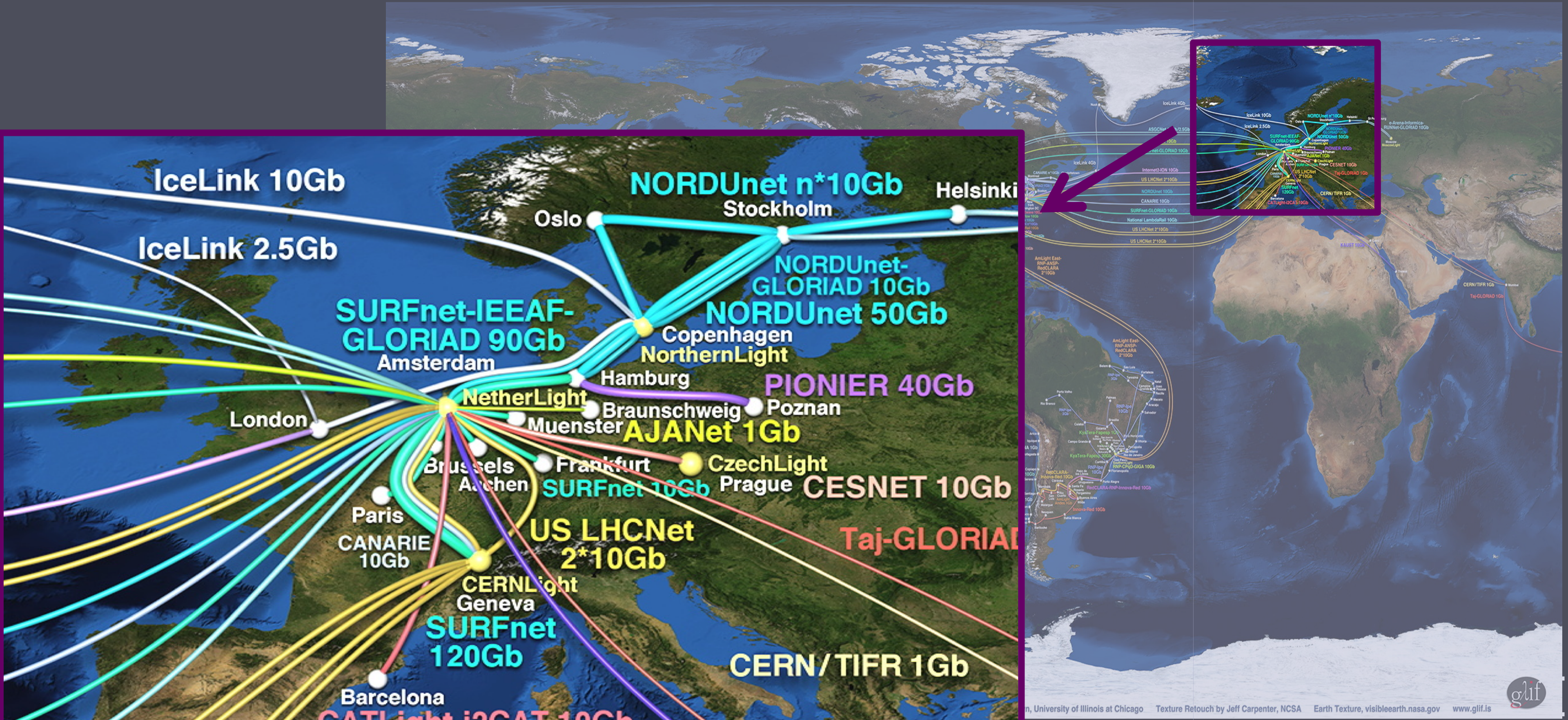
- Move towards streaming
  - Netflix
  - youtube
- Same in science world
  - SKA / LOFAR
  - Light Source
  - Environmental (Marine, Meteorology, …)
- Data is not always huge
  - Sometimes it is very complex
  - Some example:
    - biodiversity

The GLIF – LightPaths around the World

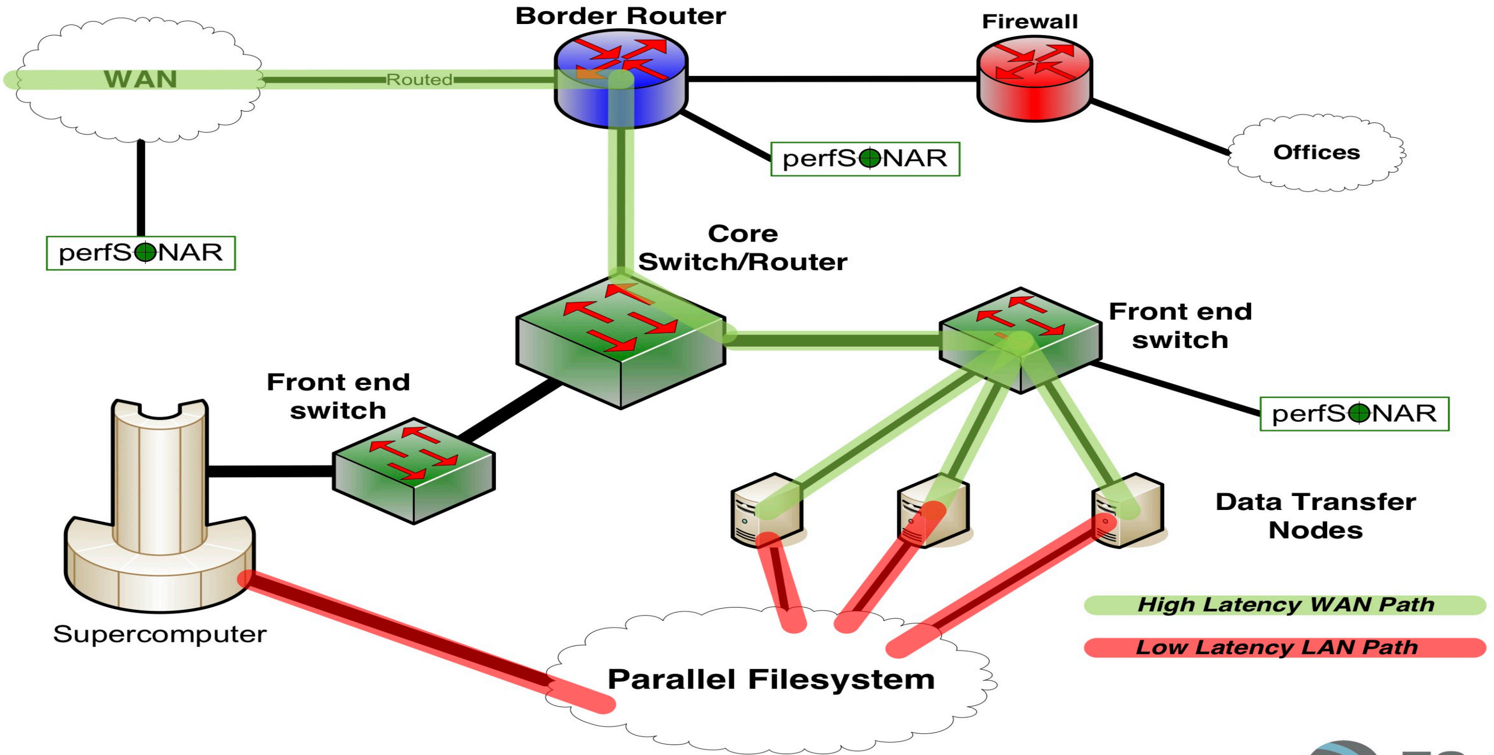


Amsterdam is a major hub in The GLIF

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.



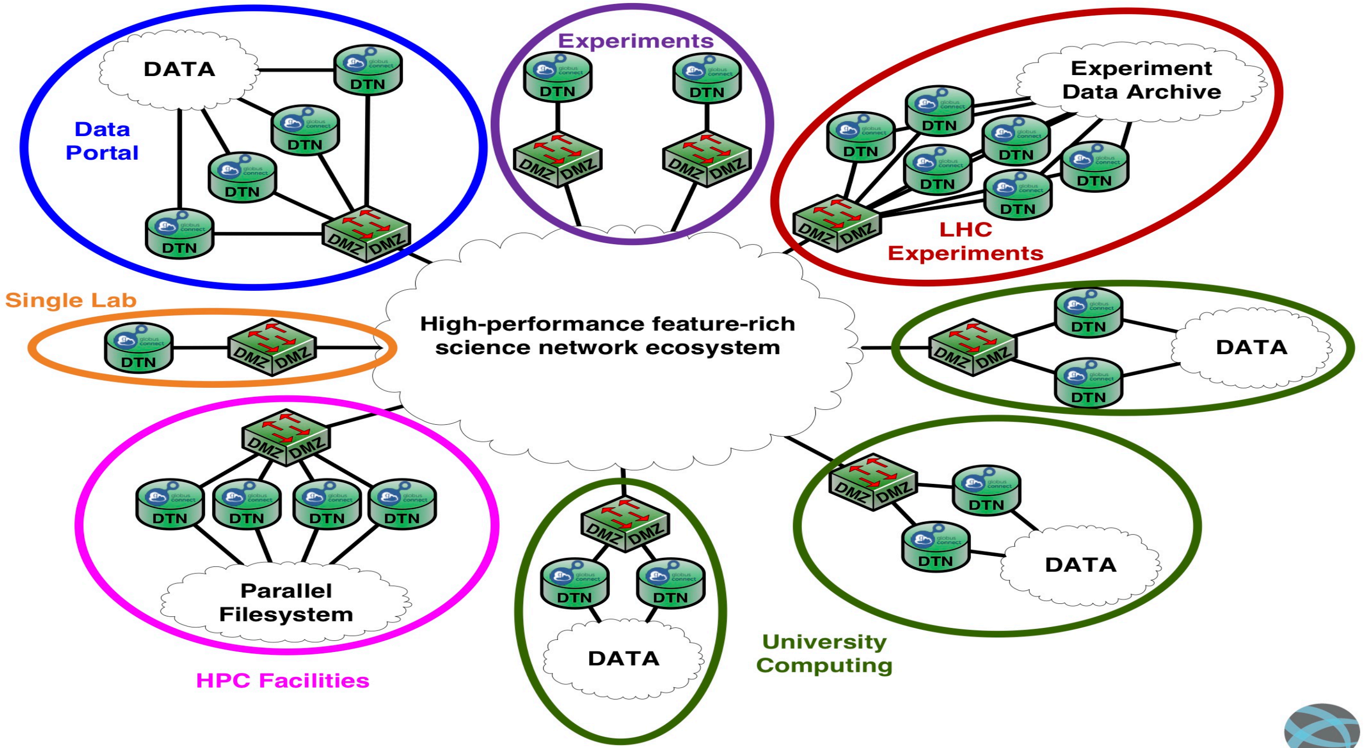
Science DMZ – HPC Center DTN Cluster



ESnet

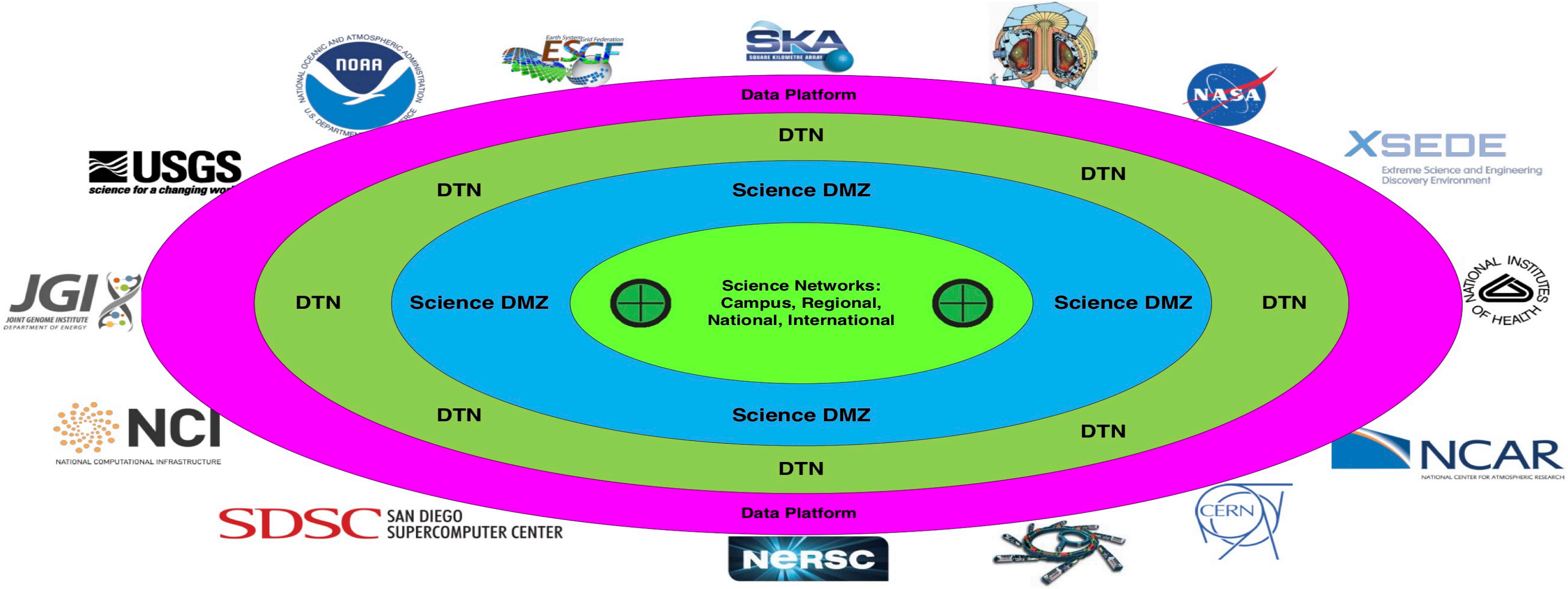
Courtesy Eli Dart, ESnet

Science DMZs for Science Applications



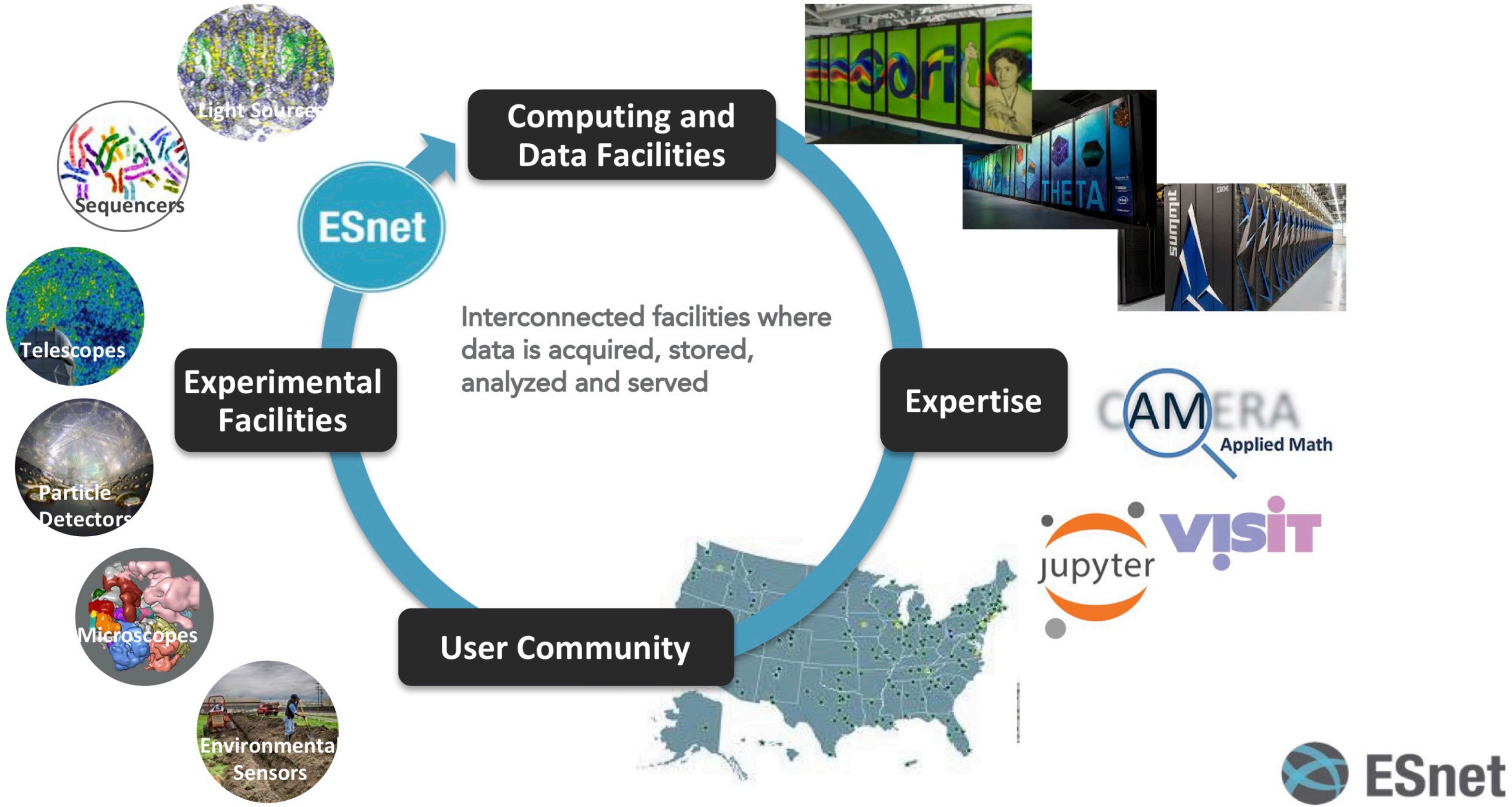
Courtesy Eli Dart, ESnet

Data Ecosystem – Concentric View



Courtesy Eli Dart, ESnet

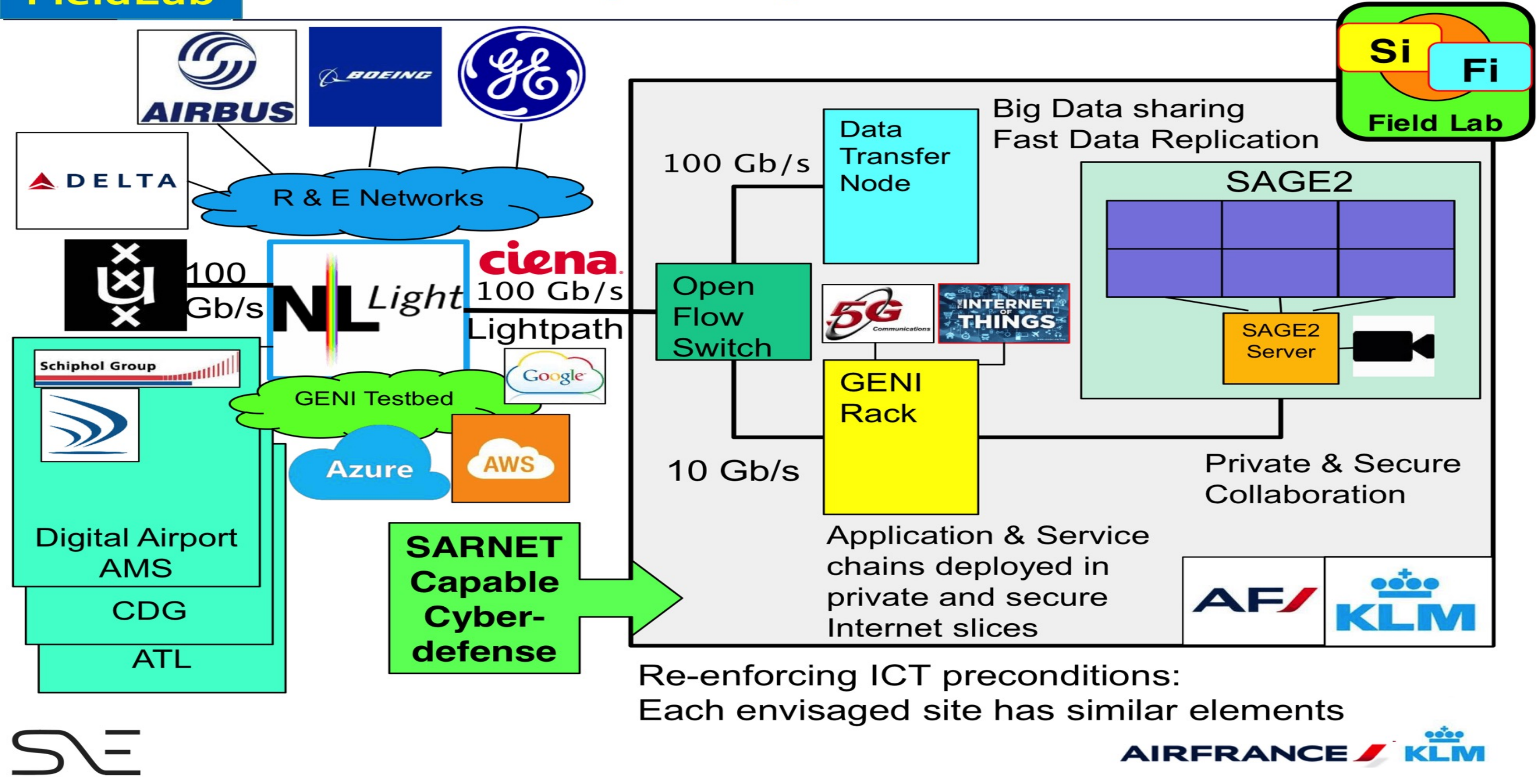
Superfacility Model for Productive, Reproducible Science



Extreme App Store for Science; Amazonizing infrastructure

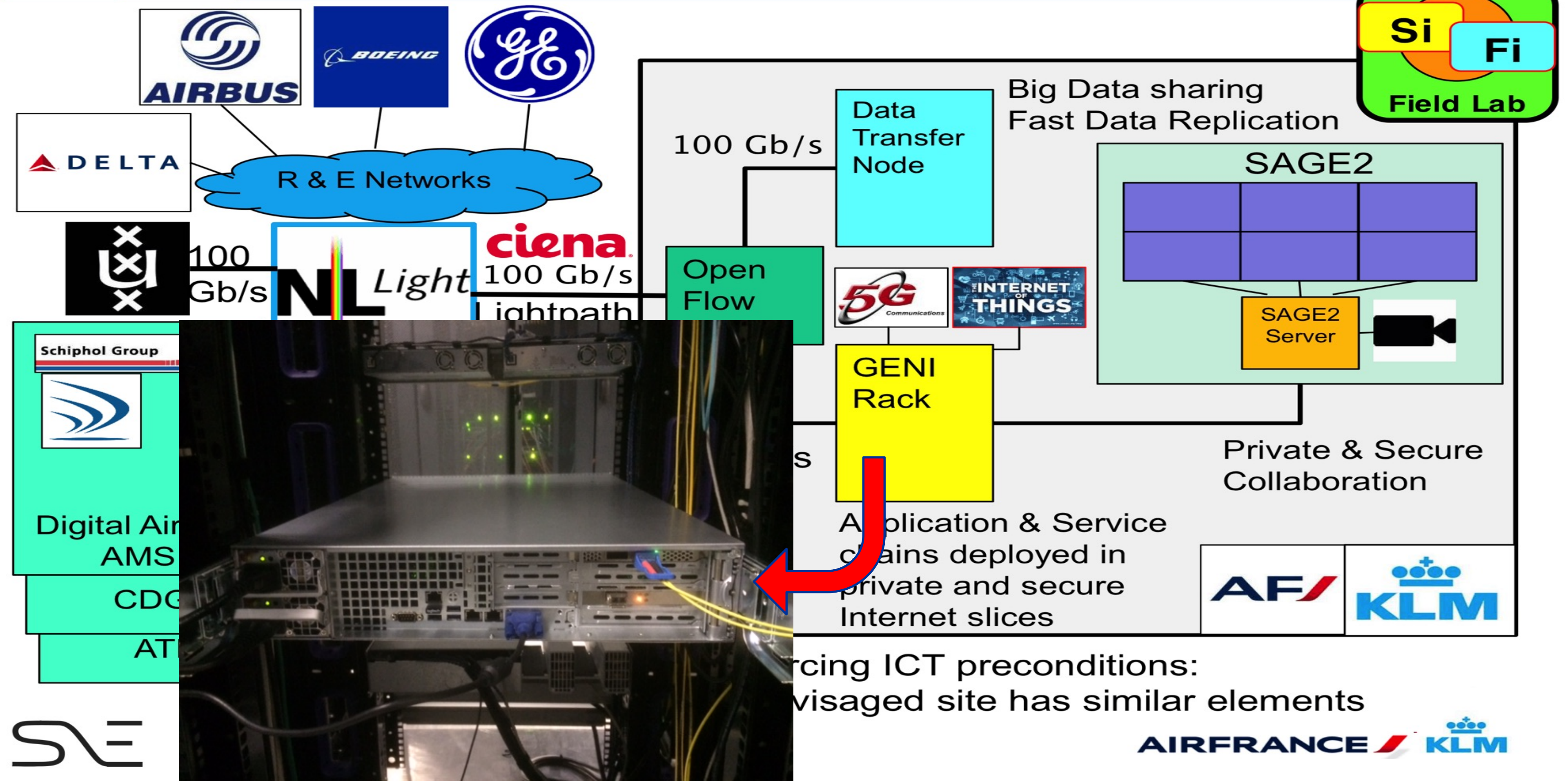
AF/KLM FieldLab

Ambition to put capabilities into fieldlab



AF/KLM FieldLab

Ambition to put capabilities into fieldlab



AF/KLM FieldLab

Ambition to p… …ab

AIRBUS | BOEING | GE | DELTA

R & E Networks

100 Gb/s

NL Light

ciena 100 Gb/s Lightpath

Si Fi Field Lab

Schiphol Group

Digital Air… AMS

CDG

AT…

GENI Rack

…s

Application & Service chains deployed in private and secure Internet slices

Private & Secure Collaboration

AF KLM

…rcing ICT preconditions:

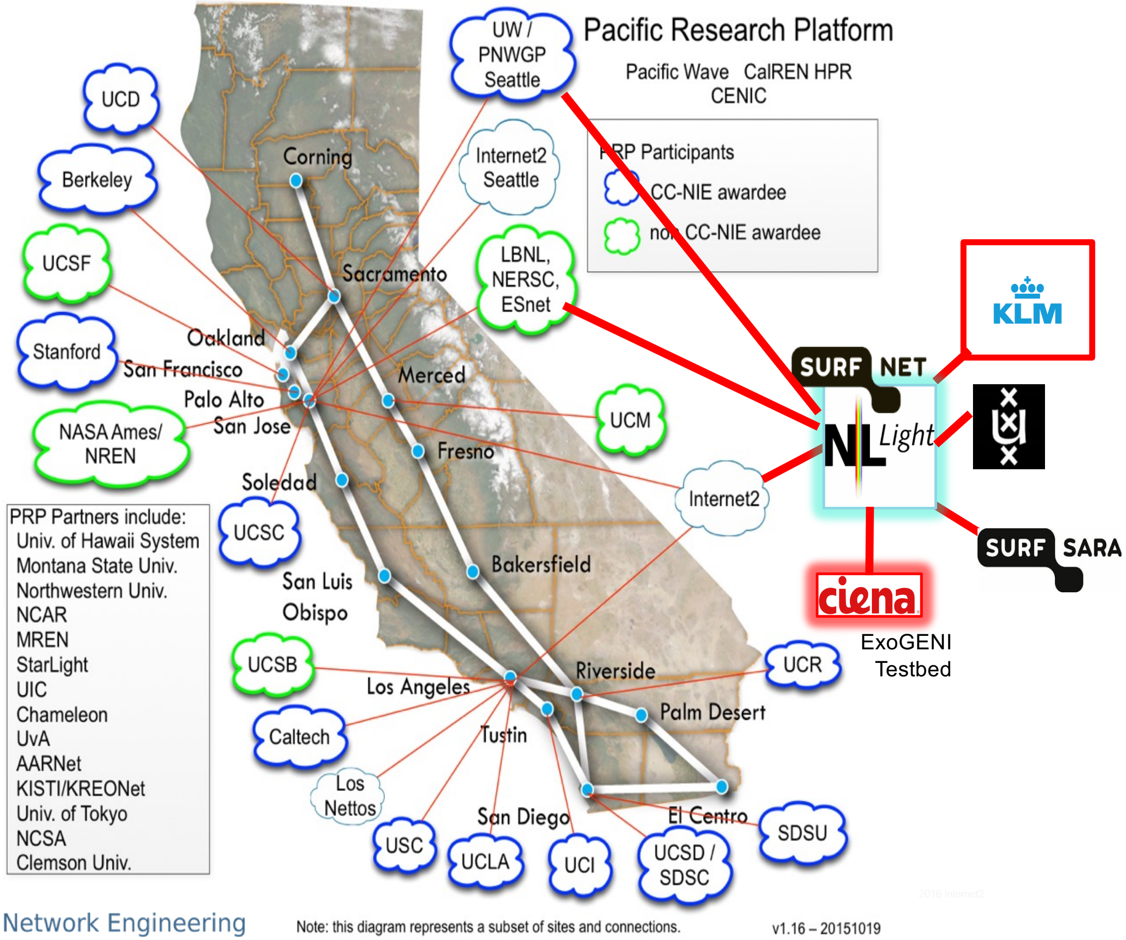
…visaged site has similar elements

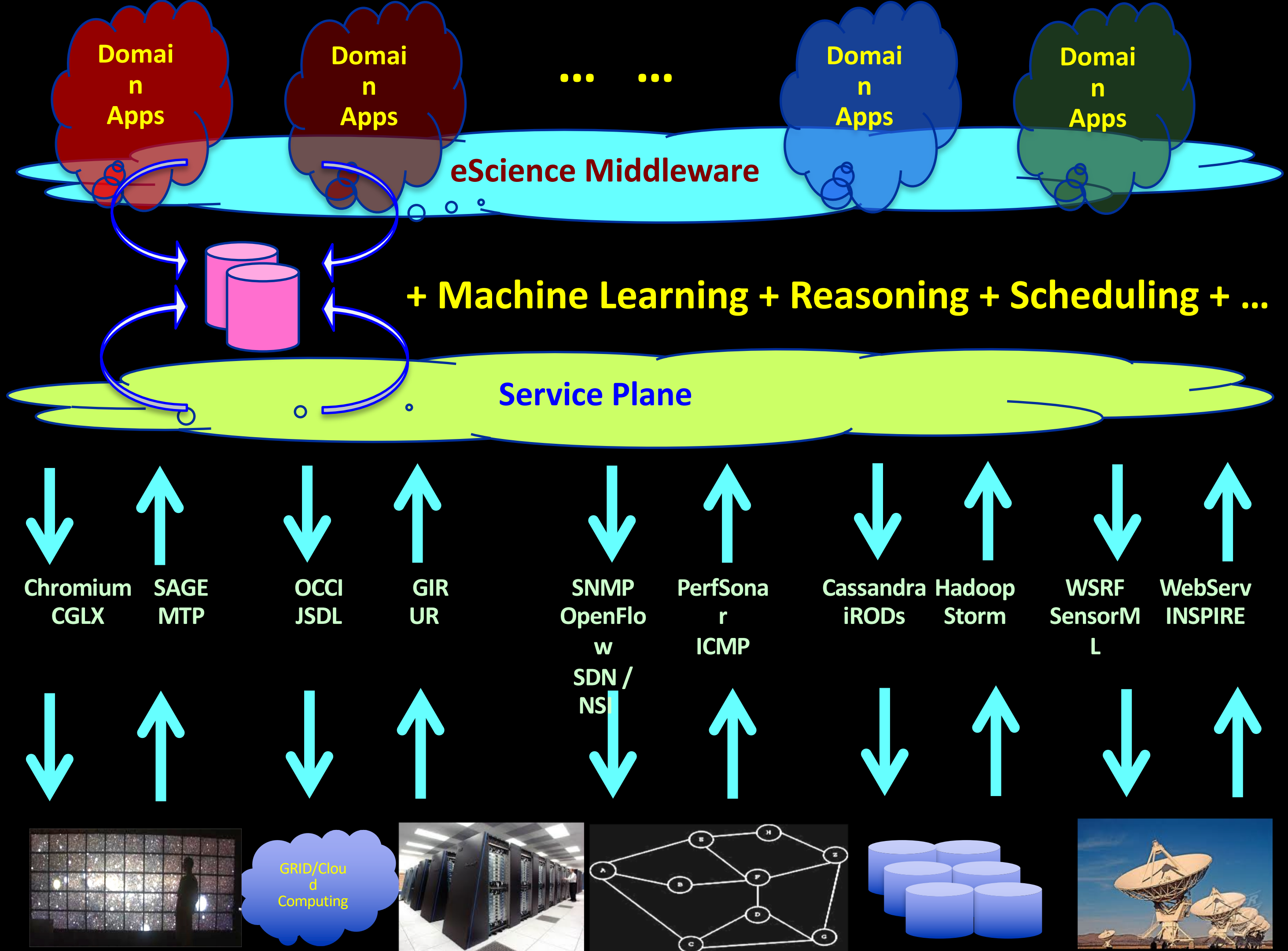
AIRFRANCE KLM

Pacific Research Platform testbed involvement

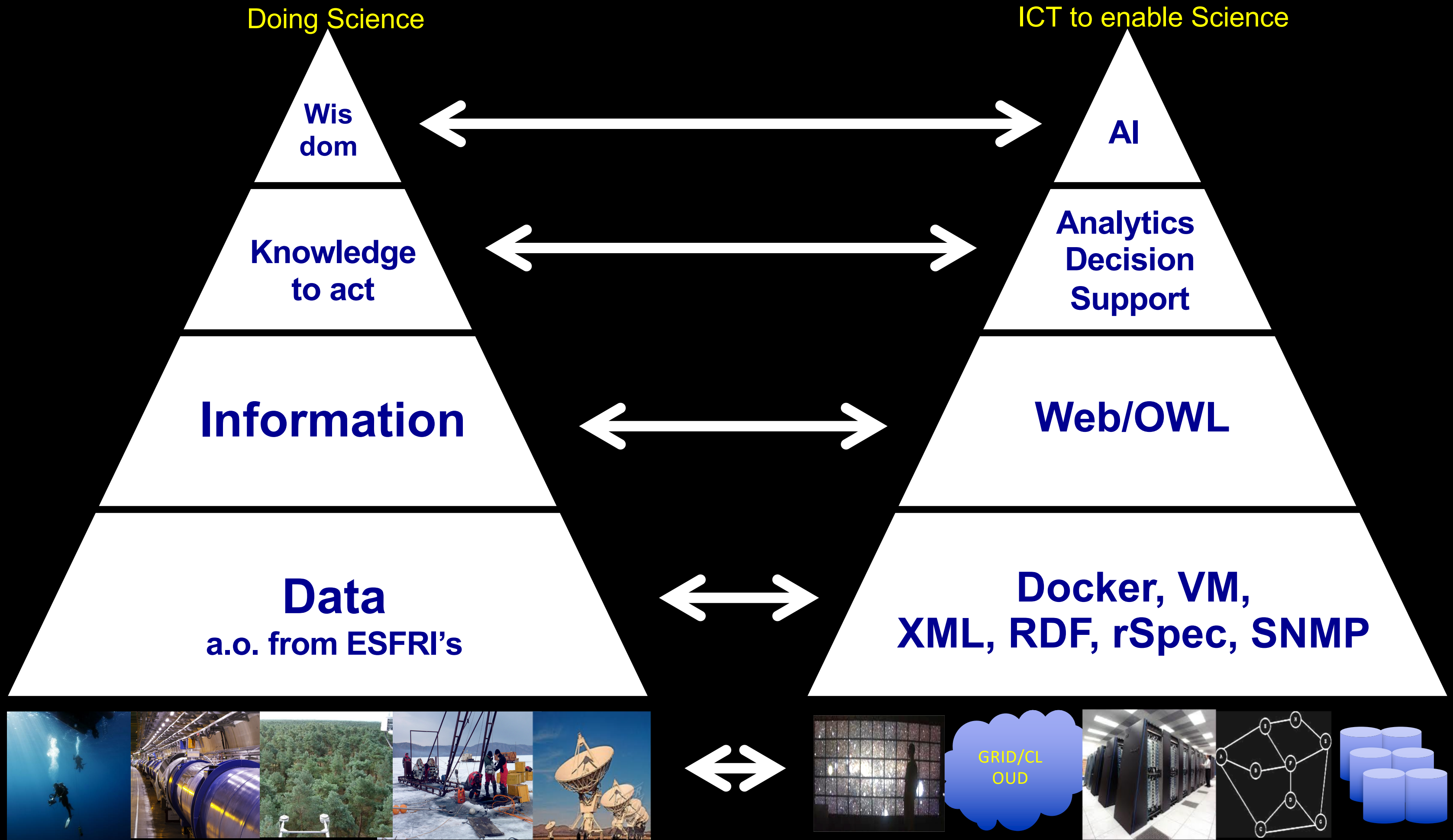
Research goal:
Explore value of academic network research capabilities that enable innovative ways & models to share big data assets

PACIFIC RESEARCH PLATFORM

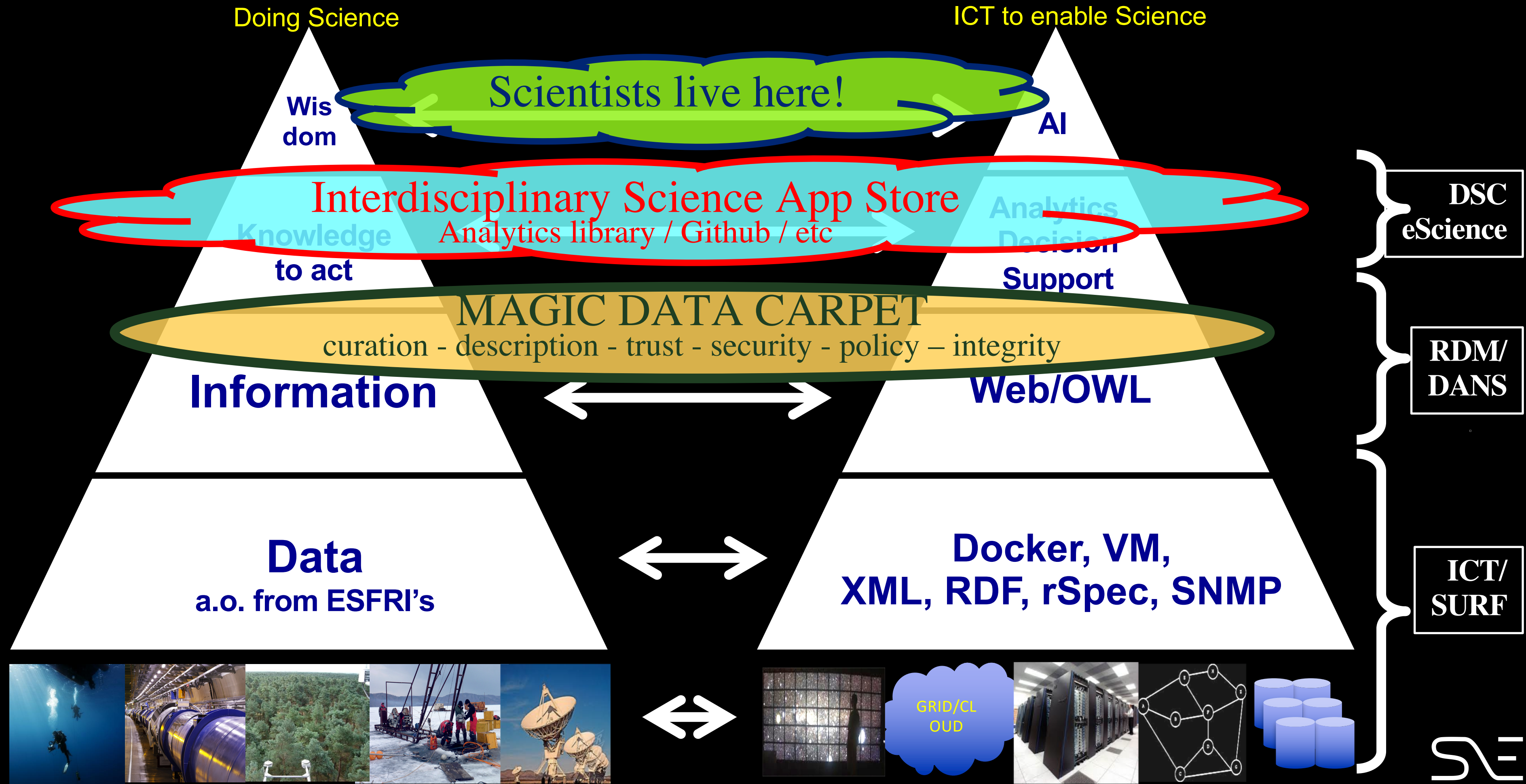



Domain Apps
Domain Apps
… …
Domain Apps
Domain Apps
eScience Middleware
+ Machine Learning + Reasoning + Scheduling + …
Service Plane
Chromium CGLX
SAGE MTP
OCCI JSDL
GIR UR
SNMP OpenFlow SDN / NS
PerfSonar ICMP
Cassandra iRODs
Hadoop Storm
WSRF SensorML
WebServ INSPIRE
GRID/Cloud Computing

The Big Data Challenge



The Big Data Challenge



SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonnais, Leon Gommans, Rodney Wilson, Lydia Meijer, Frank Fransen Tom van Engers, Paola Grosso, Gauravdeep Shami, Cees de Laat,Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

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

Lydia Meijer (PI), Cees de Laat (Co-PI), Leon Gommans, Tom van Engers, Paola Grosso, Kees Nieuwenhuis.

EPI: Enabling Personalized Interventions

Cees de Laat(PI), Sander Klous (PL), Leon Gommans, Tom van Engers, Paola Grosso, Henri Bal, Anwar Osseyran, Aki Harma, Douwe Biesma, Peter Grünwald, Floortje Scheepers, Gertjan Kaspers.

SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonnais, Leon Gommans, Rodney Wilson, Lydia Meijer, Frank Fransen Tom van Engers, Paola Grosso, Gauravdeep Shami, Cees de Laat,Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

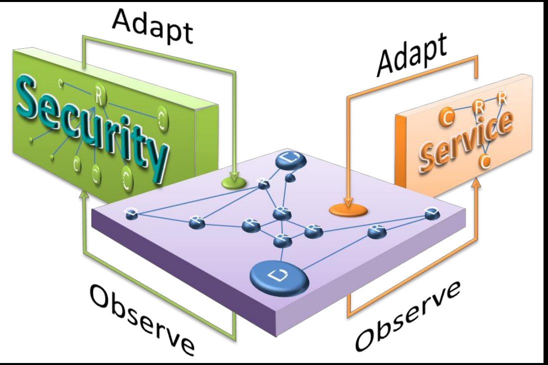
Big Data: real time ICT for logistics
Data Logistics 4 Logistics Data (dl4ld)

Lydia Meijer (PI), Cees de Laat (Co-PI), Leon Gommans, Tom van Engers, Paola Grosso, Kees Nieuwenhuis.

EPI: Enabling Personalized Interventions

Cees de Laat(PI), Sander Klous (PL), Leon Gommans, Tom van Engers, Paola Grosso, Henri Bal, Anwar Osseyran, Aki Harma, Douwe Biesma, Peter Grünwald, Floortje Scheepers, Gertjan Kaspers.

Cyber security program SARNET



Research goal is to obtain the knowledge to create ICT systems that:

- model their state (situation)
- discover by observations and reasoning if and how an attack is developing and calculate the associated risks
- have the knowledge to calculate the effect of counter measures on states and their risks
- choose and execute one.

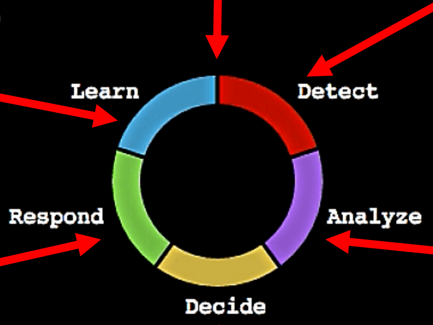
In short, we research the concept of networked computer infrastructures exhibiting SAR: Security Autonomous Response.

SARNET Publications (subset)

Laboratory: ExoGeni & PRP
Fieldlab with KLM & CIENA
OSA-Optical Forum Conference paper [1]

SC16 INDIS workshop paper [2]
TNC short paper [4]

IEEE NetSoft paper [5]
efficiency of SDN mitigations [8]



SC16 demo/poster <delaat.net/sc16>
Salt Lake City (UT)
IEEE Sec-Virtnet 2016 paper [3]
Removing Undesirable Flows [6]

SC15 demo/poster <delaat.net/sc16>
Austin (TX)
Sec-Virtnet paper [3]
Computational Trust Model (SCTM) [7]

SC16 demo/poster <delaat.net/sc16>
Salt Lake City (UT)
CoreFlow: Enriching Bro security events [9]

1. Paper: R. Koning, A. Deljoo, S. Trajanovski, B. de Graaff, P. Grosso, L. Gommans, T. van Engers, F. Fransen, R. Meijer, R. Wilson, and C. de Laat, "Enabling E-Science Applications with Dynamic Optical Networks: Secure Autonomous Response Networks ", OSA Optical Fiber Communication Conference and Exposition, 19-23 March 2017, Los Angeles, California.
2. Paper: Ralph Koning, Nick Buraglio, Cees de Laat, Paola Grosso, "CoreFlow: Enriching Bro security events using network traffic monitoring data.", Special section on high-performance networking for distributed data-intensive science, SC16", Future Generation Computer Systems, <accepted for publication>
3. Paper: Ralph Koning, Ben de Graaff, Cees de Laat, Robert Meijer, Paola Grosso, "Analysis of Software Defined Networking defenses against Distributed Denial of Service attacks", The IEEE International Workshop on Security in Virtualized Networks (Sec-VirtNet 2016) at the 2nd IEEE International Conference on Network Softwarization (NetSoft 2016), Seoul Korea, June 10, 2016.
4. Short paper: Nick Buraglio, Ralph Koning, Cees de Laat, Paola Grosso, "Enriching network and security events for event detection", Conference proceedings TNC2017, https://tnc17.geant.org/core/presentation/30.
5. Paper: Ralph Koning, Ben de Graaff, Robert Meijer, Cees de Laat, Paola Grosso, "Measuring the effectiveness of SDN mitigations against cyber attacks", IEEE Conference on Network Softwarization (Netsoft 2017 - SNS 2017), Bologna, Italy, July 3-7, 2017.
6. Paper: Gleb Polevoy, Stojan Trajanovski, Paola Grosso and Cees de Laat, "Removing Undesirable Flows by Edge Deletion.", COCOA'2018 conference, December 15 - 17, 2018, Atlanta, Georgia, USA, Springer-Verlag.
7. Paper: Ameneh Deljoo , Tom van Engers, Leon Gommans, Cees de Laat, "Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners". In: Proceedings of 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS), Dallas, TX, USA, 2018
8. Paper: R. Koning, B. de Graaff, G. Polevoy, R. Meijer, C. de Laat, P. Grosso, "Measuring the efficiency of SDN mitigations against attacks on computer infrastructures", Future Generation Computer Systems 91, 144-156.
9. Ralph Koning, Nick Buraglio, Cees de Laat, Paola Grosso, "CoreFlow: Enriching Bro security events using network traffic monitoring data.", Special section on high-performance networking for distributed data-intensive science, SC16", Future Generation Computer Systems

SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonnais, Leon Gommans, Rodney Wilson, Lydia Meijer, Frank Fransen Tom van Engers, Paola Grosso, Gauravdeep Shami, Cees de Laat,Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

ciena · UvA · NWO Netherlands Organisation for Scientific Research · TNO · AIRFRANCE KLM · COMMIT/

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

Lydia Meijer (PI), Cees de Laat (Co-PI), Leon Gommans, Tom van Engers, Paola Grosso, Kees Nieuwenhuis.

ZZ · ORACLE · TNO · UvA · ciena · AIRFRANCE KLM · Gemeente Amsterdam · evofenedex · NWO Netherlands Organisation for Scientific Research · TRANSFIDES · THALES

EPI: Enabling Personalized Interventions

Cees de Laat(PI), Sander Klous (PL), Leon Gommans, Tom van Engers, Paola Grosso, Henri Bal, Anwar Osseyran, Aki Harma, Douwe Biesma, Peter Grünwald, Floortje Scheepers, Gertjan Kaspers.

UvA · St Antonius · PHILIPS · SURF SARA · Universitair Medisch Centrum Utrecht · CWI · VU University Amsterdam · VU medisch centrum · KPMG cutting through complexity · Ministerie van Economische Zaken · ZonMw · NWO Netherlands Organisation for Scientific Research · Commit 2 Data connecting business and science

Harvard Business Review

HARVARD BUSINESS PUBLISHING

Harvard Business Review

Subscribe | Sign In | Register

ECONOMY

Managing Our Hub Economy

by Marco Iansiti and Karim R. Lakhani

FROM THE SEPTEMBER–OCTOBER 2017 ISSUE

WHAT TO READ NEXT

The IT Transformation Health Care Needs

SUMMARY SAVE SHARE 3 COMMENT TEXT SIZE PRINT $8.95 BUY COPIES

THOMAS M. SCHEER/EYEEM/GETTY IMAGES

I. The Problem

The global economy is coalescing around a few digital superpowers. We see unmistakable evidence that a winner-take-all world is emerging in which a small number of "hub firms"—including Alibaba, Alphabet/Google, Amazon, Apple, Baidu, Facebook, Microsoft, and Tencent—occupy central positions. While creating real value for users, these companies are also capturing a disproportionate and expanding share of the value, and that's shaping our collective economic future. The very same technologies that promised to democratize business are now threatening to make it more monopolistic.

Data value creation monopolies

➔

Create an equal playing field

➔

Sound Market principles

https://hbr.org/2017/09/managing-our-hub-economy

Data Sharing: 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 or Algorithms may have to be processed in foreign data centers.
  - How to organize alliances?
  - How to translate from strategic via tactical to operational level?
  - How to enforce policy using modern Cyber Infrastructure?
  - What are the different fundamental data infrastructure models to consider?

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

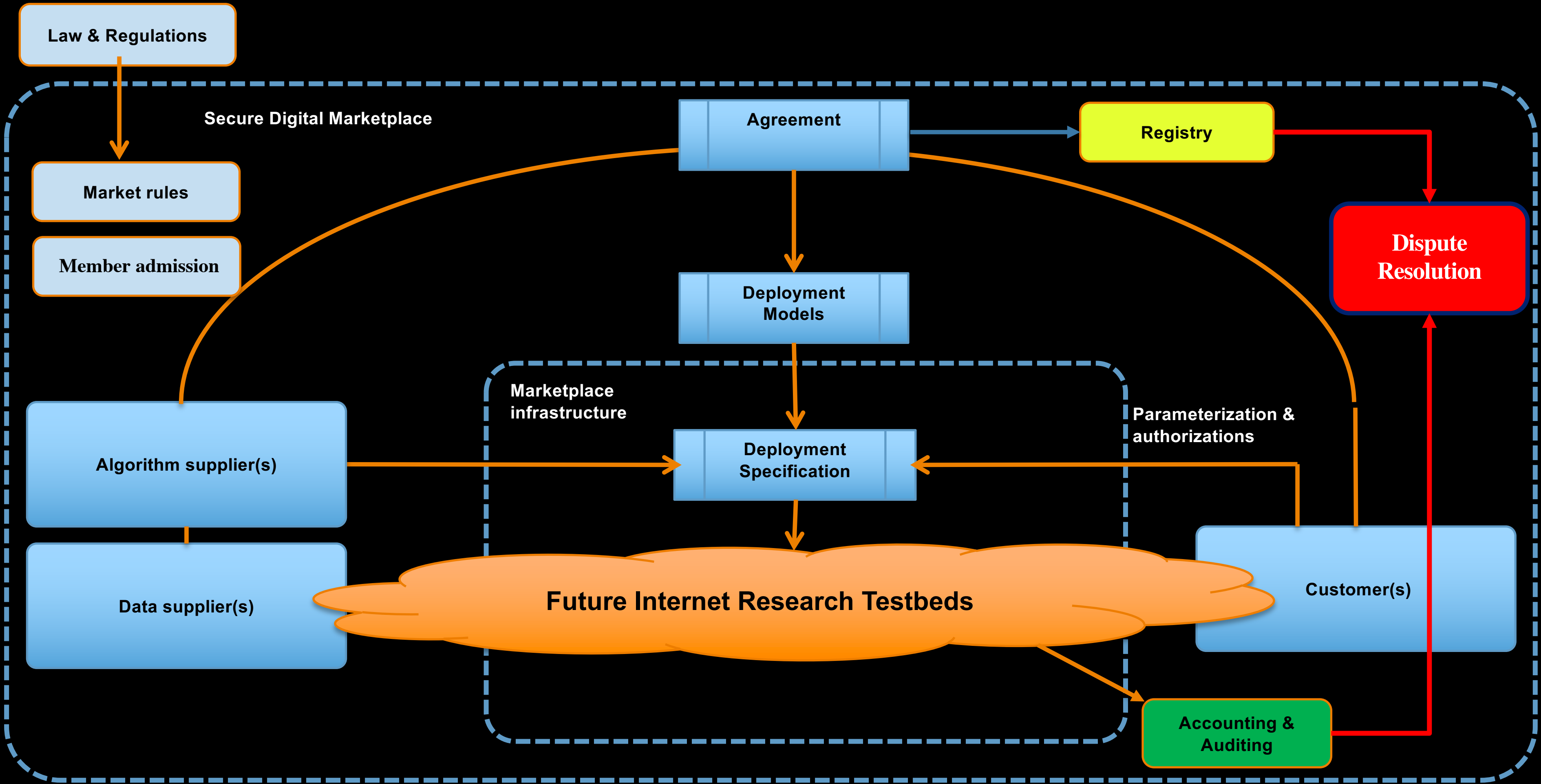
Cargo Logistics Data
(C1) DaL4LoD
(C2) Secure scalable policy-enforced distributed data Processing
(using blockchain)

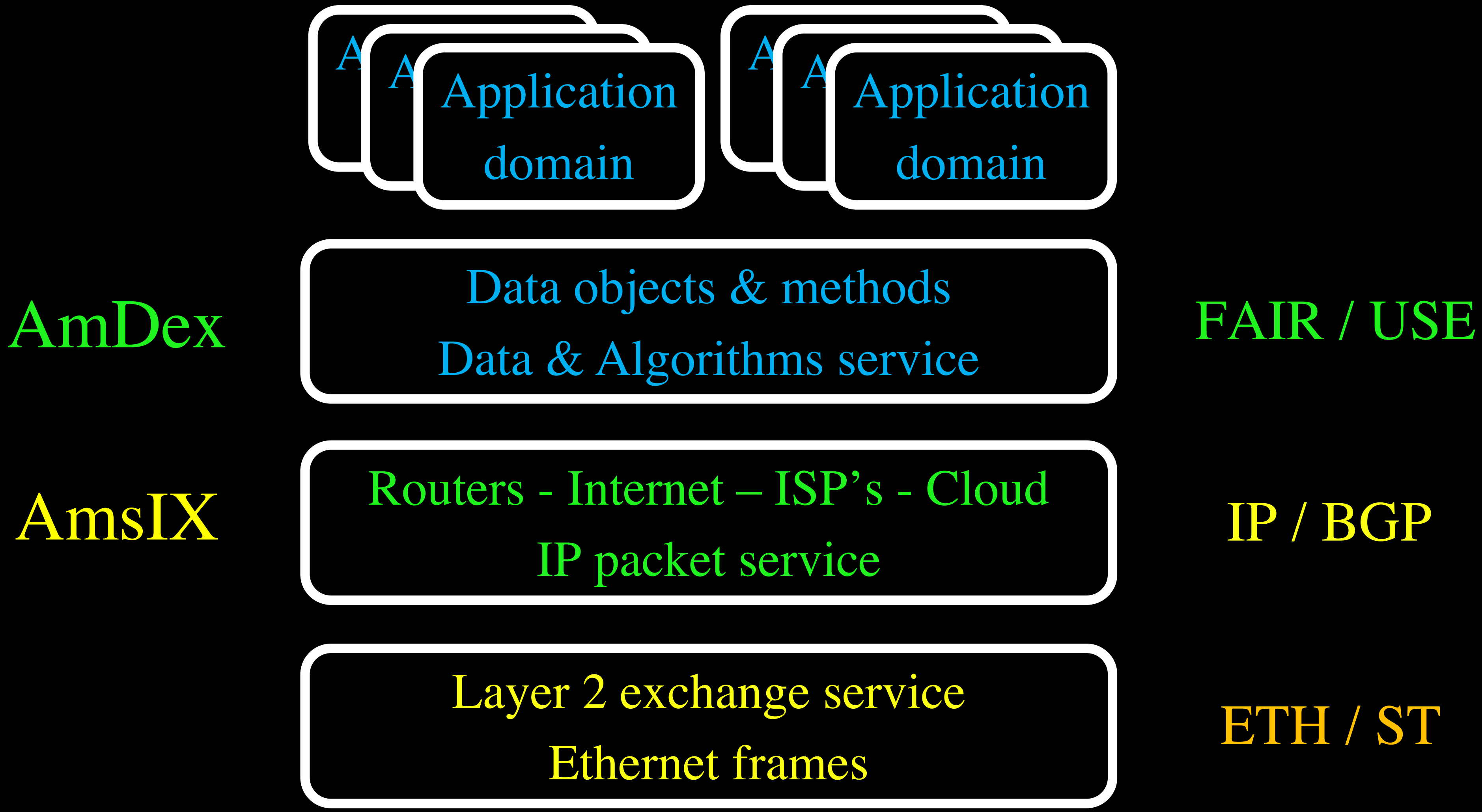
NLIP iShare project

Cybersecurity Big Data
NWO COMMIT/
SARNET project
3.5 FTE

iSHARE
powered by NLIP

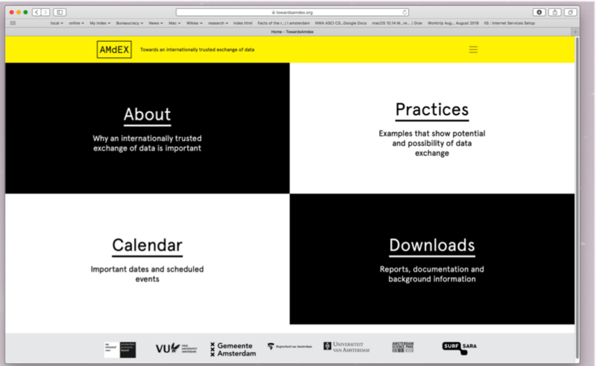
Secure Digital Market Place Research

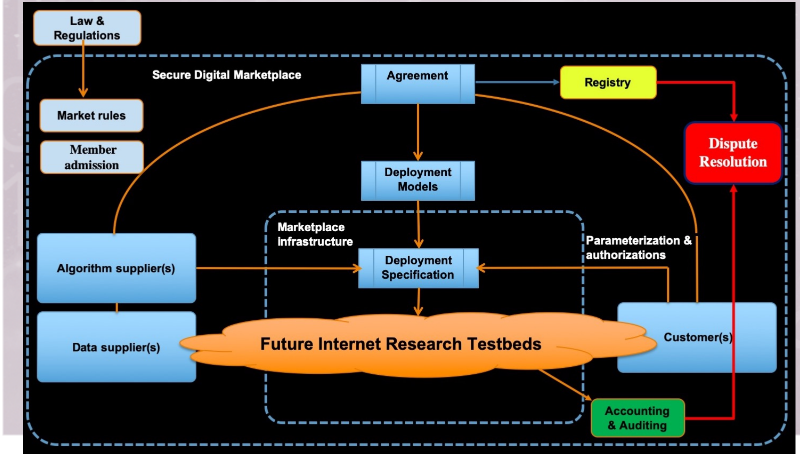



A
A
Application domain
A
A
Application domain
AmDex
Data objects & methods
Data & Algorithms service
FAIR / USE
AmsIX
Routers - Internet – ISP’s - Cloud
IP packet service
IP / BGP
Layer 2 exchange service
Ethernet frames
ETH / ST

AMdEX.eu

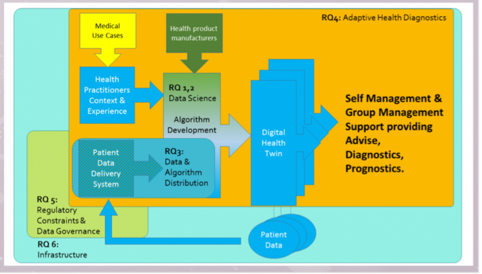
- Competing organisations, share data for common benefit
- Trust, Risk, data ownership & control
  - Industry: AF-KLM, Health, etc
  - Science: European Open Science Cloud
  - Society: Amsterdam Economic Board





Aircraft Maintenance AF-KLM

Health: Enabling Personal Interventions



SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonnais, Leon Gommans, Rodney Wilson, Lydia Meijer, Frank Fransen Tom van Engers, Paola Grosso, Gauravdeep Shami, Cees de Laat,Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

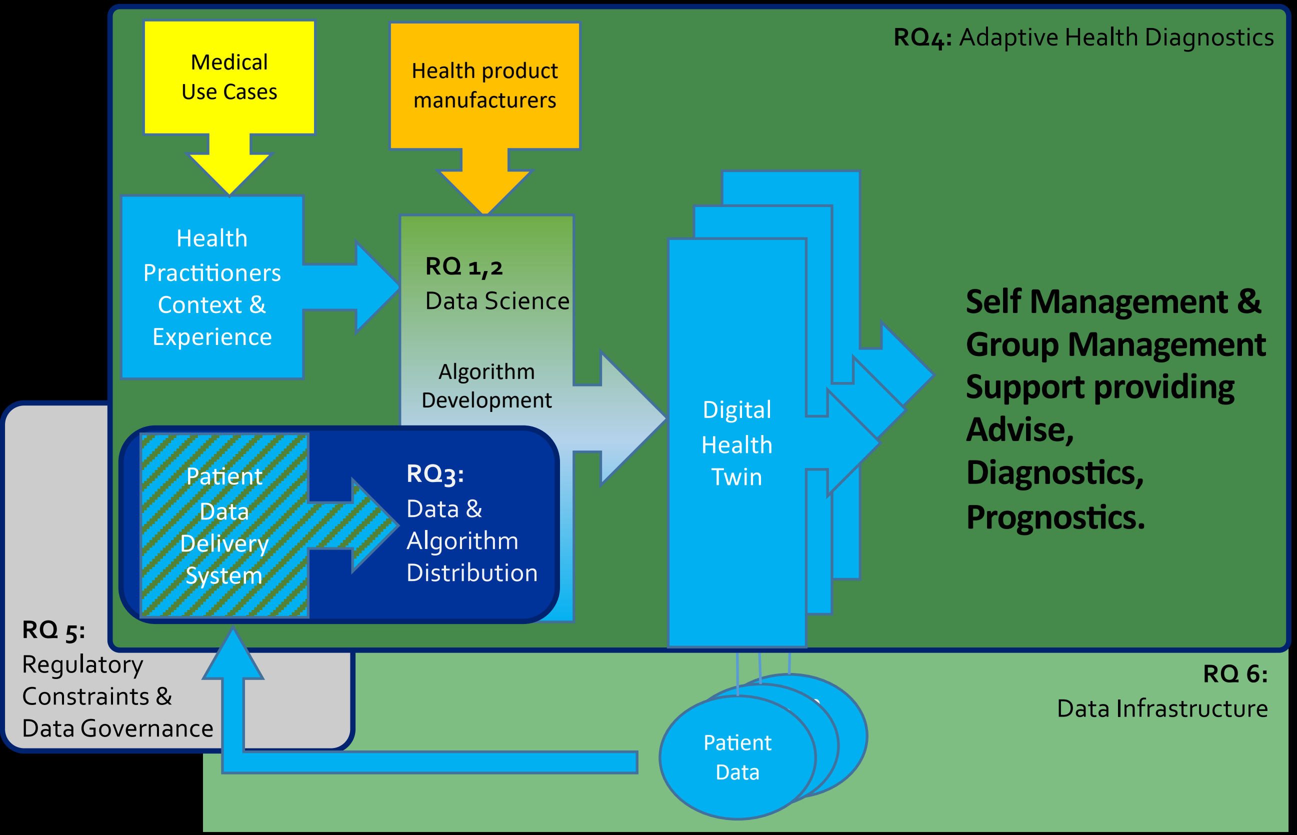
Big Data: real time ICT for logistics Data Logistics 4 Logistics Data (dl4ld)

Lydia Meijer (PI), Cees de Laat (Co-PI), Leon Gommans, Tom van Engers, Paola Grosso, Kees Nieuwenhuis.

EPI: Enabling Personalized Interventions

Cees de Laat(PI), Sander Klous (PL), Leon Gommans, Tom van Engers, Paola Grosso, Henri Bal, Anwar Osseyran, Aki Harma, Douwe Biesma, Peter Grünwald, Floortje Scheepers, Gertjan Kaspers.

Health Use Case
Enabling Personal Interventions



https://delaat.net/epi

The overall aim of this project is to explore the use and effectiveness of data driven development of scientific algorithms, supporting personalized self- and joint management during medical interventions / treatments.

The key objective is to use data science promoting health practically with data from various sources to formulate lifestyle advice, prevention, diagnostics, and treatment tailored to the individual and to provide personalized effective real-time feedback via a concept referred in this proposal as a digital health twin.

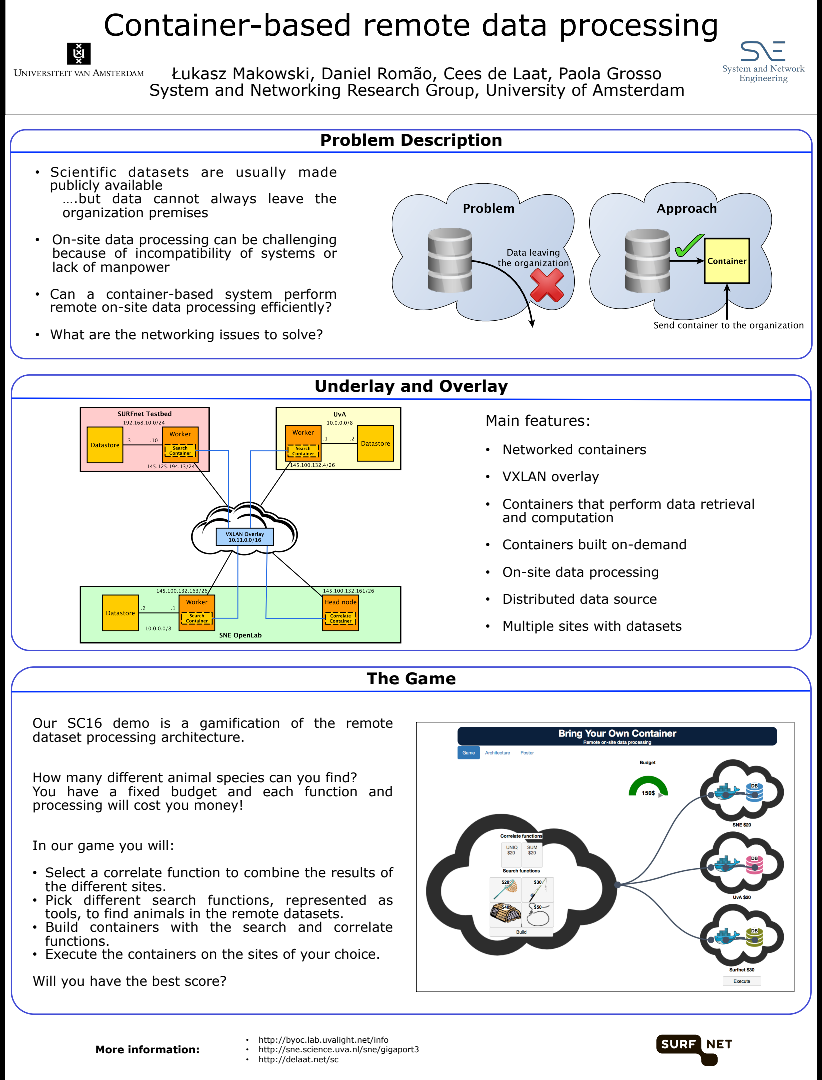
Research questions

- RQ1: Dynamically Analyzing Interventions based on Small Groups: how can we determine, based on as little data as possible, whether an intervention does or does not work for a small group or even an individual patient?
- RQ2: Dynamically Personalizing the Group: how can we identify effective intervention strategies and optimize personalization strategies applicable for different patient and lifestyle profiles via dynamic (on-line) clustering of patients? Can those clusters be adapted as new data about patients and results of interventions come in and as other data may be removed or modified?
- RQ3: Data and Algorithm Distribution: what are the consequences of a distributed, multi-platform, multi-domain, multi-data-source big data infrastructure on the machine learning algorithms and what are potential consequences on performance?
- RQ4: Adaptive health diagnosis leading to optimized intervention: how can we enhance self- / joint management by dynamically integrating updated models generated from machine learning from various data sources in state of the art health support systems that based on personal health records, knowledge of health modes and effective interventions?
- RQ5: Regulatory constraints and data governance: how can we create scalable solutions that meet legal requirements and consent or medical necessity-based access to data for allowed data processing and preventing breaches of these rules by embedded compliance, providing evidence trails and transparency, thus building trust in a sensitive big data sharing infrastructure?
- RQ6: Infrastructure: how can the various requirements from the use-cases be implemented using a single functional ICT-infrastructure architecture?

SC16 Demo

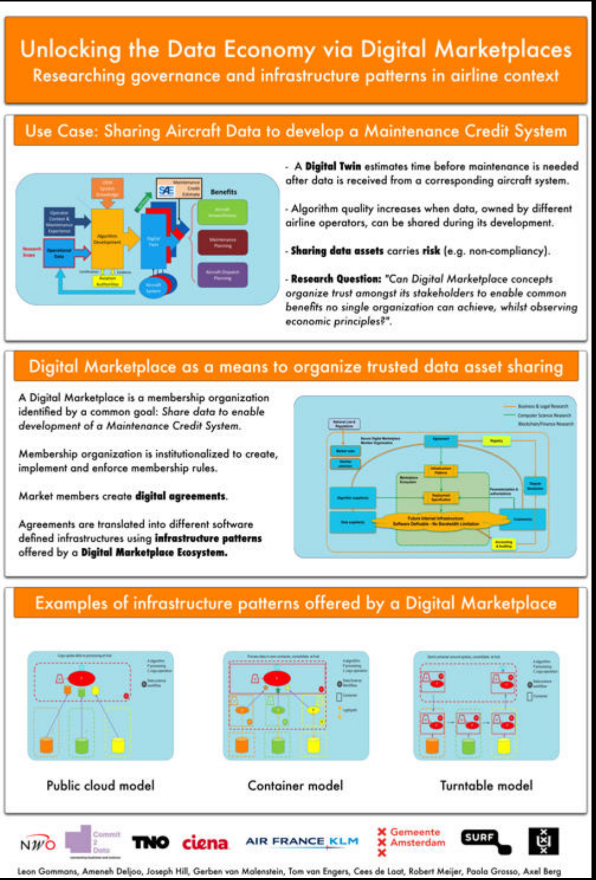
DockerMon
Sending docker containers with search algorithms to databases all over the world.

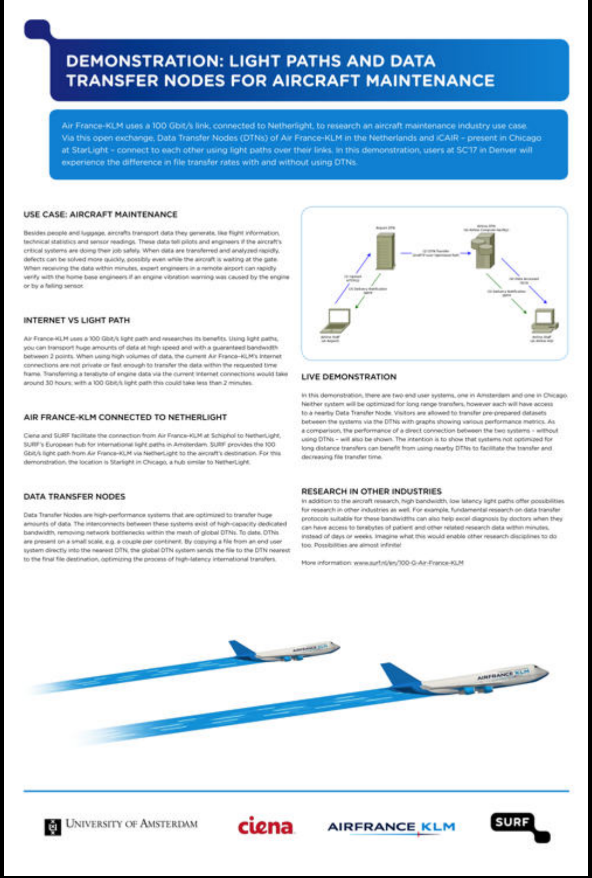
http://sc.delaat.net/sc16/index.html#5

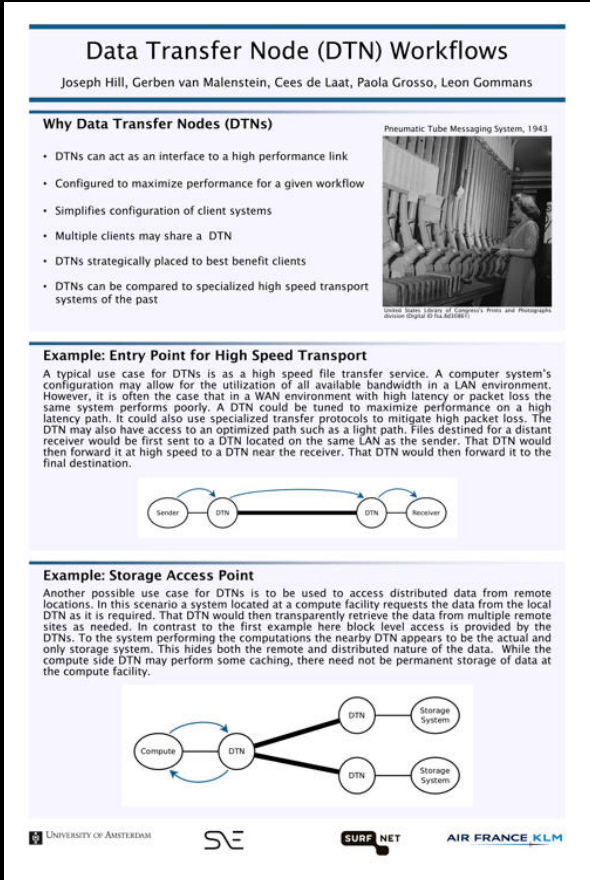


SC17 Posters and proof of concepts & demo's

http://sc.delaat.net/sc17

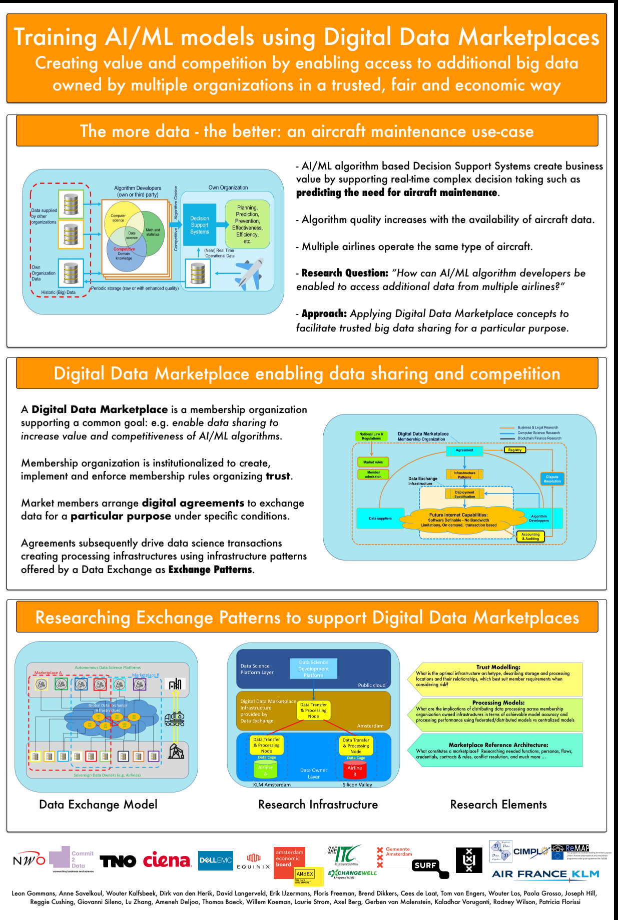


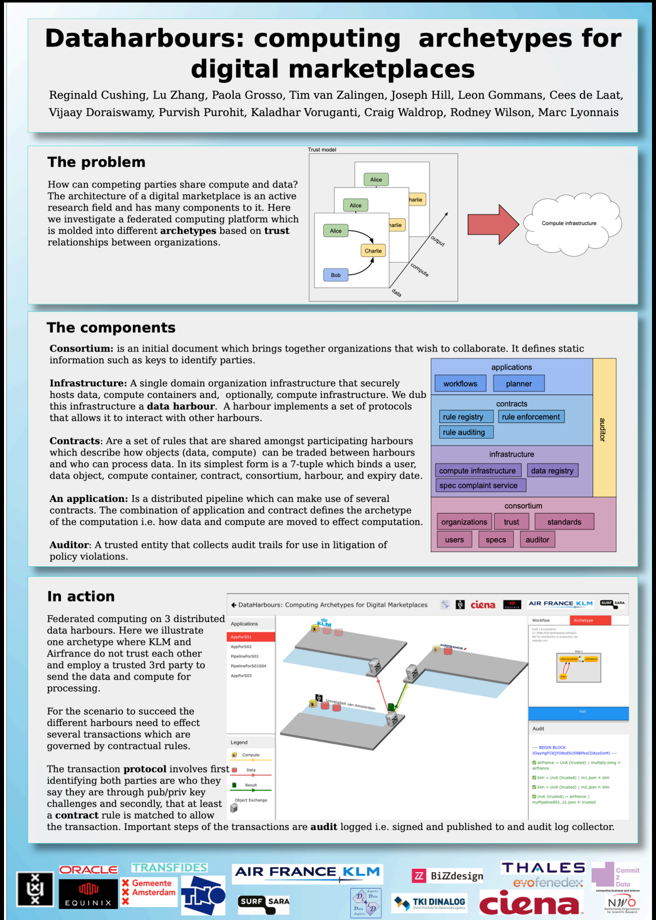




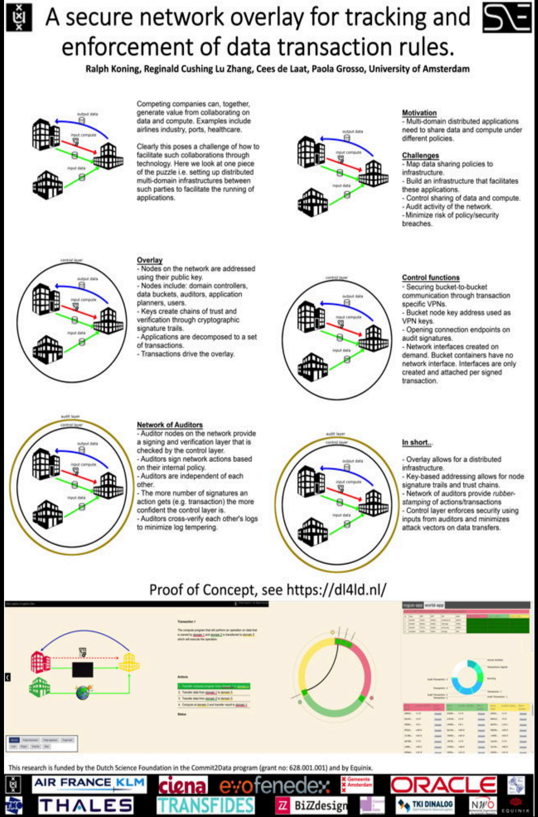
SC18 – Dallas TX

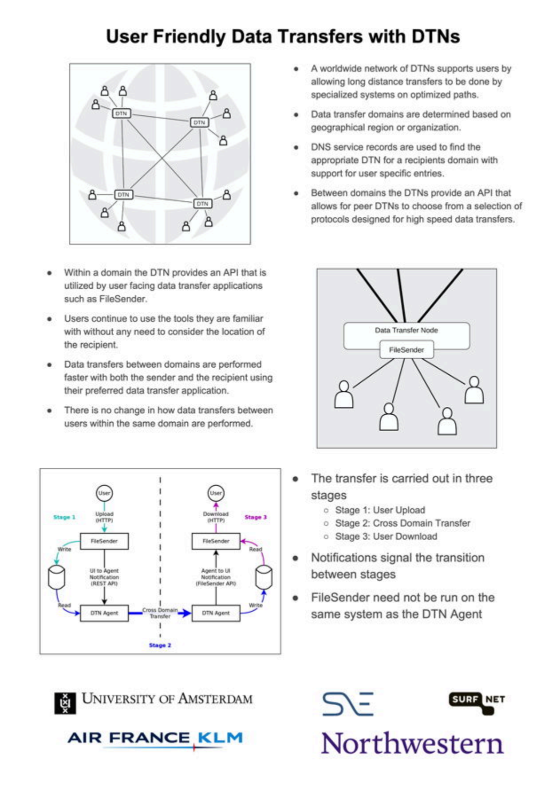
http://sc.delaat.net

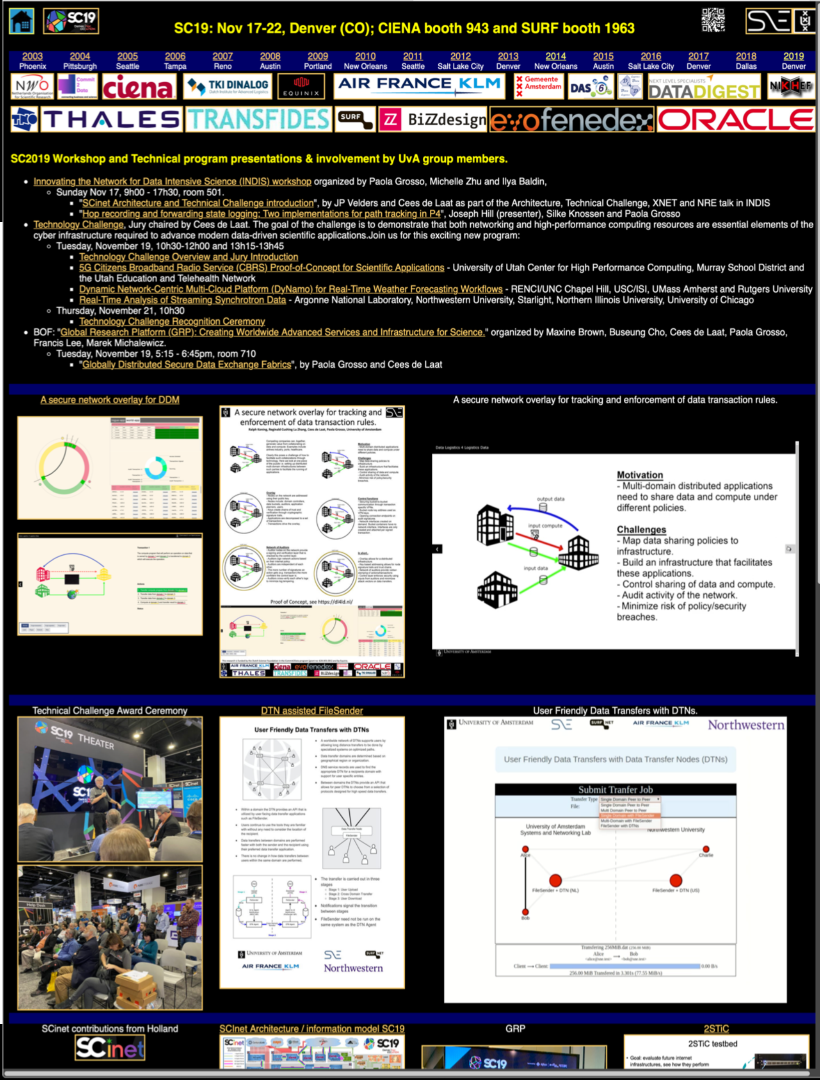




SC19 – Denver CO
https://sc.delaat.net







A secure network overlay for monitoring and enforcement of multi-domain applications.

Conclusions, Info, Acknowledgements, Q&A

- Data hindered by risk of unexpected use, lack of trust
- Using market principles, enforcement and determining incentives and value in the data life cycle to make data flow
- More information:
  - http://delaat.net/dl4ld http://delaat.net/epi
  - https://www.esciencecenter.nl/project/secconnet
  - https://towardsamdex.org