ICT to support the transformation of Science in the Roaring Twenties

Cees de Laat

Systems and Networking Laboratory
Complex Cyber Infrastructure group
University of Amsterdam
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.
Transformations

- Internet
- Computing
- Data
- Science
Transformations

- Internet
  - From end to end to client server bubbles

- Computing

- Data

- Science
Fading Trust in Internet

Dependency
Trust

Research Gap!

1980

2017
Internet Backbone
The GLIF – LightPaths around the World
A cable landing station may or may not be required, depending on whether, for example, the submarine cable requires power to power submarine repeaters or amplifiers. The voltages applied to the cables can be high **3,000 to 4,000 volts** for a typical trans-Atlantic telecommunications cable system, and 1,000 volts for a cross-channel telecommunications cable system. Submarine power cables can operate at many kilovolts: for example, the [Fenno-Skan power cable operates at 400 kV DC](https://www.underseacables.com/cable-systems/Fenno-Skan).
Per fiber: ~ 80-100 colors * 50 - 100 GHz
Per color: 10 – 40 – 100 - 200 – 400 Gbit/s
Max total: ~20 Tbit/s = ~2 Tbyte/s

New: Hollow Fiber!
⇒ less RTT!
It is a bit freaky with this wireless technology.

protocol LAN due to the easy comparison and convenience in the digital home. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.
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 support...
Internet moves from IXP’s into datacenters

LIMITED TIME OFFERS

Bundles & Fragmentation

THE FILTER BUBBLE
What the Internet Is Hiding from You
ELI PARISER

FIRST MONTH FREE SUBSCRIPTION
The Trend

• Internet used to be end user to end user or service
  – Meshed network
  – Internet exchanges
  – Net Neutrality

• It is becoming end user to data center
  – Internal data center “meet me” rooms
  – Data centers interconnect based on business
  – Less and less data via Internet exchanges
  – Neutrality may get violated by filtering, policing

• And we are back where we started, a bundled phone system.
Transformations

- Internet
  - From end to end to client server bubbles
- Computing
  - From Dinosaurs to Ant Colonies
- Data
- Science
Some progress

2018

= ~7×

? 540 MHz
? MFlops
1000 MByte memory
16000 MByte ssd
0.0012 kWh – 18 h

1976

80 MHz
160 MFlops
8 MByte memory
300 MByte disks
120 kW
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
Internet of Things

Cloud

Services

Creative Strategies, Inc
Transformations

• Internet
  – From end to end to client server bubbles

• Computing
  – From Dinosaurs to Ant Colonies

• Data
  – From putting Data in the Cloud to getting it back

• Science
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
Data value creation monopolies

Create an equal playing field

Sound Market principles

Science DMZs for Science Applications

High-performance feature-rich science network ecosystem

- Data Portal
- Experiments
- Experiment Data Archive
- LHC Experiments
- Single Lab
- Parallel Filesystem
- HPC Facilities
- University Computing

Courtesy Eli Dart, ESnet
Data Ecosystem – Concentric View
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?
Secure Digital Market Place Research

- Law & Regulations
- Market rules
- Member admission
- Secure Digital Marketplace
  - Agreement
  - Deployment Models
  - Deployment Specification
  - Parameterization & authorizations
  - Accounting & Auditing
- Future Internet Research Testbeds
  - Customer(s)
  - Dispute Resolution
  - Registry
  - Algorithm supplier(s)
  - Data supplier(s)
  - Market rules infrastructure
  - Member admission
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)

- **City / regional Scale**
  - NLIP iShare project

- **Campus / Enterprise Scale**
  - Cybersecurity Big Data
    - NWO COMMIT/SARNET project
    - 3.5 FTE
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
Transformations

- Internet
  - From end to end to client server bubbles
- Computing
  - From Dinosaurs to Ant Colonies
- Data
  - From putting Data in the Cloud to getting it back
- Science
  - Pulling it all together
In most applications, utilization of Big Data often needs to be combined with Scalable Computing.

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
Fire Modeling Workflows in WIFIRE

- Real-time sensors
- PRP
- Weather forecast
- Monitoring & fire mapping
- Fire perimeter
- Landscape data

SDSC SAN DIEGO SUPERCOMPUTER CENTER
UC San Diego
One Piece of the Puzzle: Vegetation Classification using Satellite Imagery
Scientific progress will be driven by
- Massive data: sensors, simulations, networks
- Predictive models and adaptive algorithms
- Heterogeneous high-performance computing

Trend: Human-AI collaborations will transform the way science is done.

Human-AI insights enabled via scientific method, experimentation, & AI reinforcement learning.

Workshop report: https://www.osti.gov/biblio/1478744
The Big Data Challenge

- Doing Science
  - Wisdom
  - Knowledge to act
  - Information
  - Data a.o. from ESFRI's

- ICT to enable Science
  - AI
  - Analytics Decision Support
  - Web/OWL
  - Docker, VM, XML, RDF, rSpec, SNMP
The Big Data Challenge

Doing Science

ICT to enable Science

Wisdom

Knowledge
to act

Scientists live here!

Interdisciplinary Science App Store

Analytics library / Github / etc

MAGIC DATA CARPET

curation - description - trust - security - policy – integrity

Data
a.o. from ESFRI's

Web/OWL

Docker, VM, XML, RDF, rSpec, SNMP

Information

DSC
eScience

RDM/
DANS

ICT/
SURF

Scientists live here!
Why?

Because we can!
Questions?

http://delaat.net
http://sne.science.uva.nl
http://www.os3.nl/
http://sne.science.uva.nl/openlab/
http://pire.opensciencedatacloud.org
http://staff.science.uva.nl/~delaat/pire/
https://rd-alliance.org
http://envri.eu