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
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.
Transformations

- Internet
- Computing
- Data
- Science
TimeLine
Network programmability and virtualisation

Compute
Assembler
Fortran
C
RPC
C++
MPI
GRID
Cloud
DATA
Apps
IOT
NDN
AI
NDN
SDN
NFV
OpenFlow & p4
WIFI
WWW
TCP
IP
Ethernet

Networks
Compute
Assembler
Fortran
C
RPC
C++
MPI
GRID
Cloud
DATA
Apps
IOT
NDN
AI
NDN
SDN
NFV
OpenFlow & p4
WIFI
WWW
TCP
IP
Ethernet
Exascale and increasingly complex science applications are exponentially raising demands from underlying DOE networks, such as traffic management, operation, and security by 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 in advance and be faster, efficient, and more proactive, anticipating bottlenecks before they happen. However, managing 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-

AI creeping in the control plane

Self driving cars, so why not self driving CI?
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.
Internet moves from IXP’s into datacenters

Rapidly loosing internet transparency

A Responsible Internet to Increase Trust in the Digital World

Fading Trust in Internet

Dependency
Trust

1980

2017

Research Gap!
Some progress

2018

= \sim 7 \times \quad 1976

540 MHz
\sim 1 \text{ GFlops}
1000 \text{ MByte memory}
16000 \text{ MByte ssd}
0.0012 \text{ kWh} – 18 \text{ h}

80 MHz
160 \text{ MFlops}
8 \text{ MByte memory}
300 \text{ MByte disks}
120 kW
If my watch is several times SDSC first computer, what will I be wearing on my wrist in 35 years?
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
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
- Often it is complexity
- Some example:
  - biodiversity
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.**

**EXEMPLARS OF SCIENTIFIC ACHIEVEMENT**

<table>
<thead>
<tr>
<th>Cosmic Microwave Background</th>
<th>DNA Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Cosmic Microwave Background" /></td>
<td><img src="image2" alt="DNA Structure" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Periodic Table of the Elements</th>
<th>Special Relativity</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Periodic Table" /></td>
<td><img src="image4" alt="Special Relativity" /></td>
</tr>
</tbody>
</table>

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


U.S. DEPARTMENT OF ENERGY

Office of Science

DOE Applied Mathematics Research Program

Scientific Machine Learning Workshop (January 2018)

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

Data a.o. from ESFRI's

MAGIC DATA CARPET
- Curation
- Description
- Trust
- Security
- Policy
- Integrity

Doing Science

Scientists live here!

Interdisciplinary Science App Store
- Analytics library / Github / AI / etc

AI

Knowledge
to act

Web/OWL

Docker, VM, XML, RDF, rSpec, SNMP

ICT to enable Science

ICT

Library
2.0

ICT

Library
2.0
<table>
<thead>
<tr>
<th>Different ways of using and sharing data</th>
<th>‘Gated community’</th>
<th>Open market</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRADITIONAL FARM</td>
<td>Platform limited</td>
<td>Free choice.</td>
</tr>
<tr>
<td>• Individual self-resourcing</td>
<td>Forced shopping. No services from others.</td>
<td>Purchase from any service provider.</td>
</tr>
<tr>
<td>• How most organisations do it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FARMER’S MARKET</td>
<td>Only within the platform. No interoperabiliteit with other platforms.</td>
<td>Peer-to-peer transactions and platform independent.</td>
</tr>
<tr>
<td>• Market, sharing and exchange</td>
<td>No exclusieve control on sovereign data.</td>
<td>Full control on sovereign data.</td>
</tr>
<tr>
<td>• Social networks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data services

Data transactions

Data control
Data value creation monopolies

Create an equal playing field

Sound Market principles

AMDEX
Layer 2 exchange service
Ethernet frames
Routers - Internet – ISP’s - Cloud
IP packet service
Layer 2 exchange service
Ethernet frames

AMS-IX
Data objects & methods
Data & Algorithms service
Routers - Internet – ISP’s - Cloud
IP packet service

FAIR / USE
IP / BGP
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
The Roaring Twenties!

- In the 90’s the Internet was running on top of the telco’s.
- We freed it in the 2000’s with GLIF and the *Lights.
- We developed the computer science for virtualization of CI.
- Networking is (almost) not the problem anymore (DMC2022...).
- Data and algorithms & apps and services are now in the cloud.
- Just a few large players emerge with an almost monopoly.
- Roaring 20’s to free the Data with initiatives such as GRP!
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