

Digital Data Markets:
Trusted Data Processing in
Untrusted Environments

Cees de Laat

Systems and Networking Laboratory

University of Amsterdam

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 may have to be processed in untrusted data centers.
 - How to enforce that using modern Cyber Infrastructure?
 - How to organize such alliances?
 - How to translate from strategic via tactical to operational level?
 - What are the different fundamental data infrastructure models to consider?

Harvard Business Review



Harvard Business Review

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 COMMENT 3 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

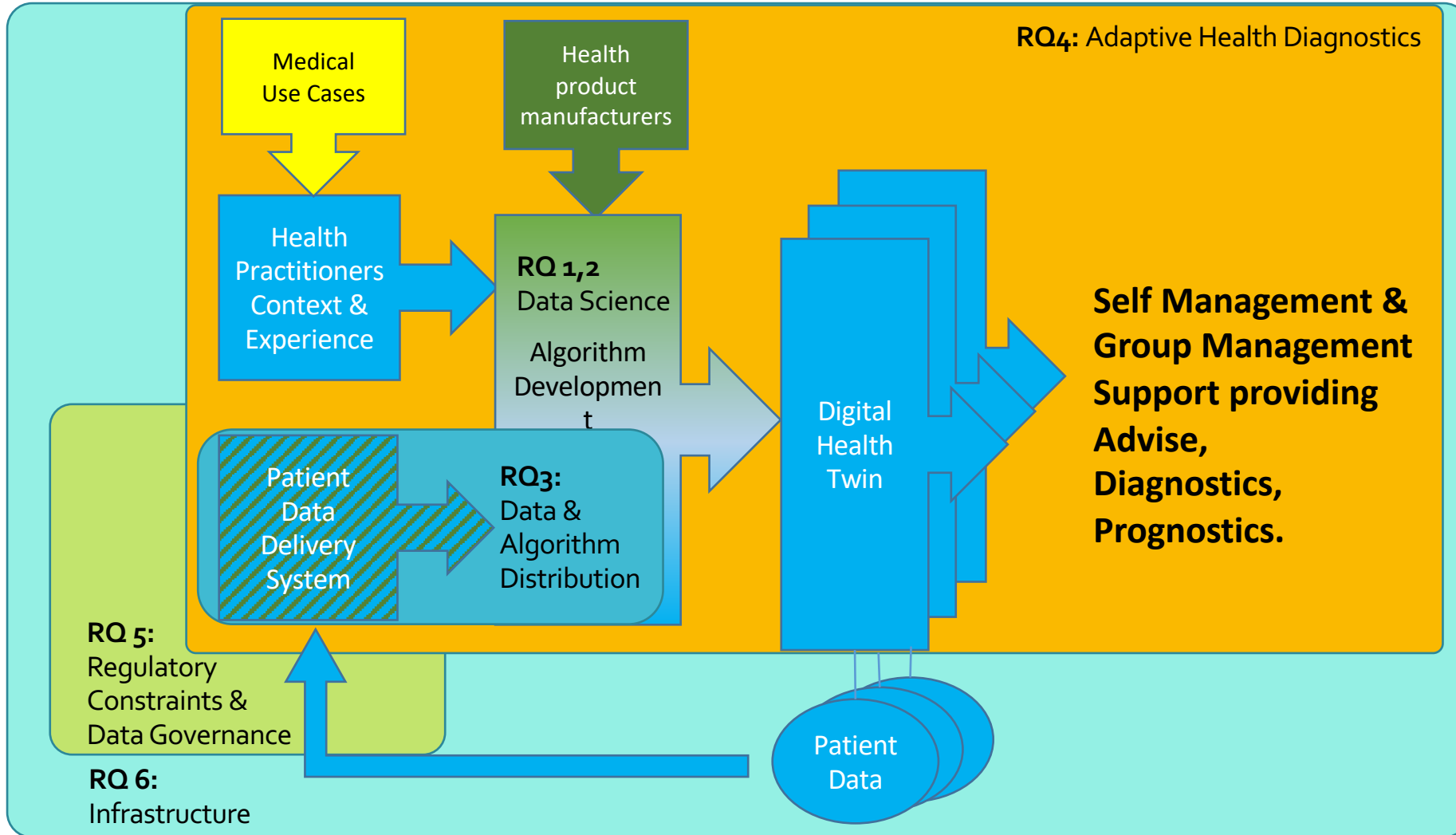


Sound Market
principles

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

Health use case

Enabling Personal Interventions



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) DL4LD
(C2) **Secure scalable policy-enforced distributed data Processing**
(using blockchain)



Cybersecurity Big Data
NWO COMMIT/
SARNET project
3.5 FTE

City / regional Scale

Campus / Enterprise Scale

NLIP iShare project



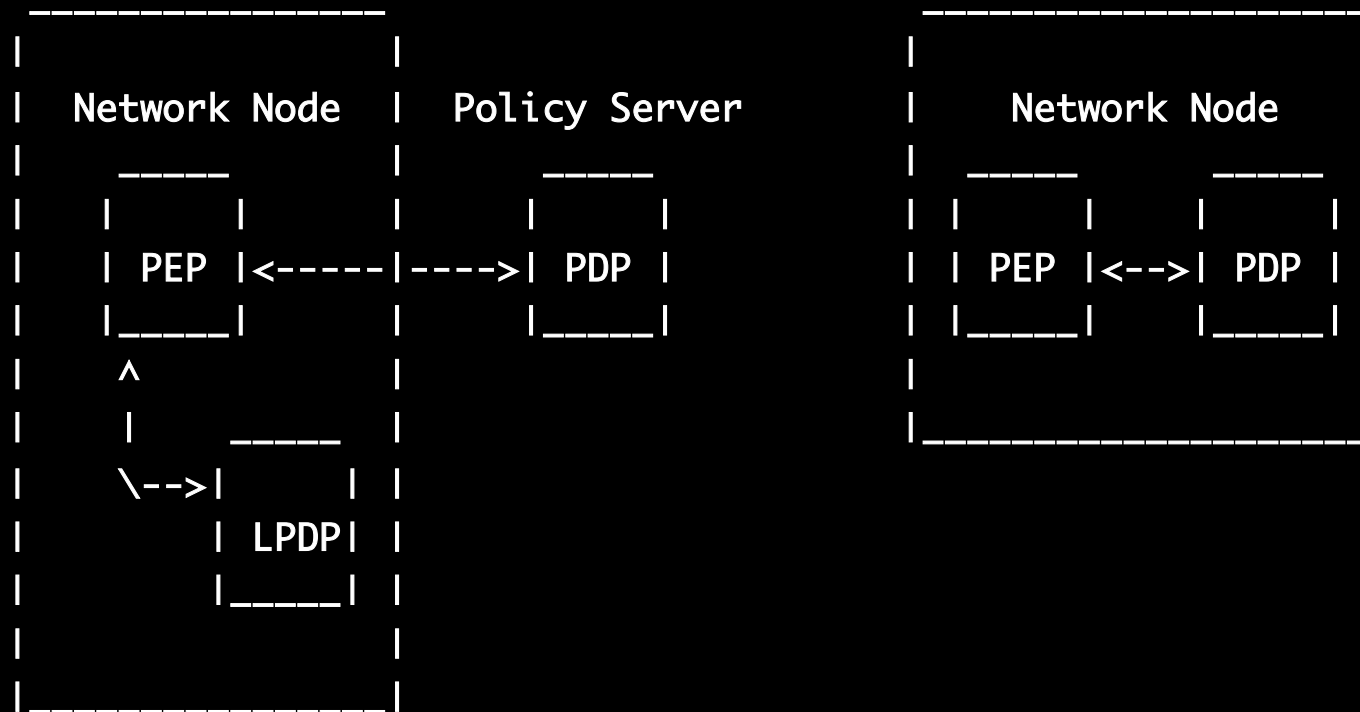
Approach

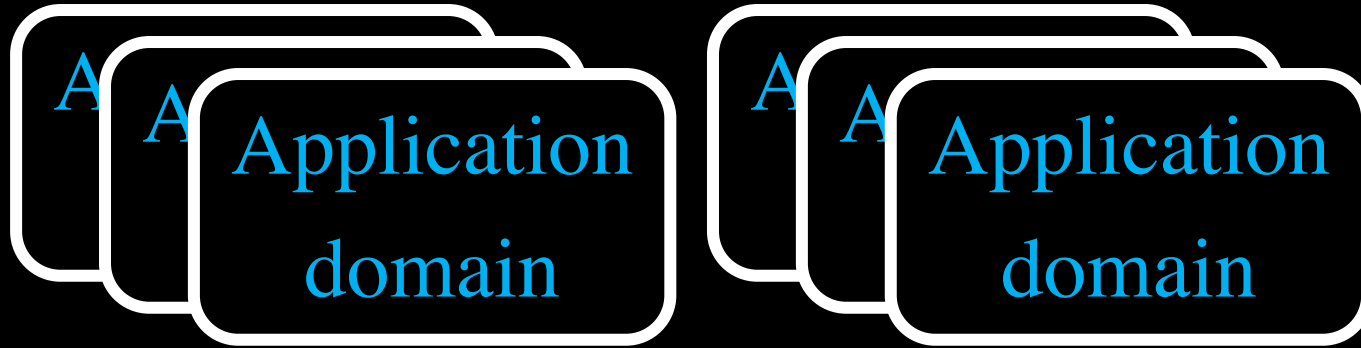
- Strategic:
 - Translate legislation into machine readable policy
 - Define data use policy
 - Trust evaluation models & metrics
- Tactical:
 - Map app given rules & policy & data and resources
 - Bring computing and data to (un)trusted third party
 - Resilience
- Operational:
 - TPM & Encryption schemes to protect & sign
 - Policy evaluation & docker implementations
 - Use VM and SDI/SDN technology to enforce
 - Block chain to record what happened (after the fact!)



IETF: Common Open Policy Service (COPS)

- Rfc 2748, 2753, 4261





AmDex

Data objects & methods
Data & Algorithms service

FAIR / USE

AmsIX

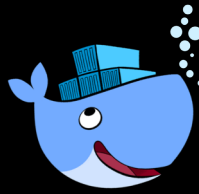
Routers - Internet – ISP's - Cloud
IP packet service

IP / BGP

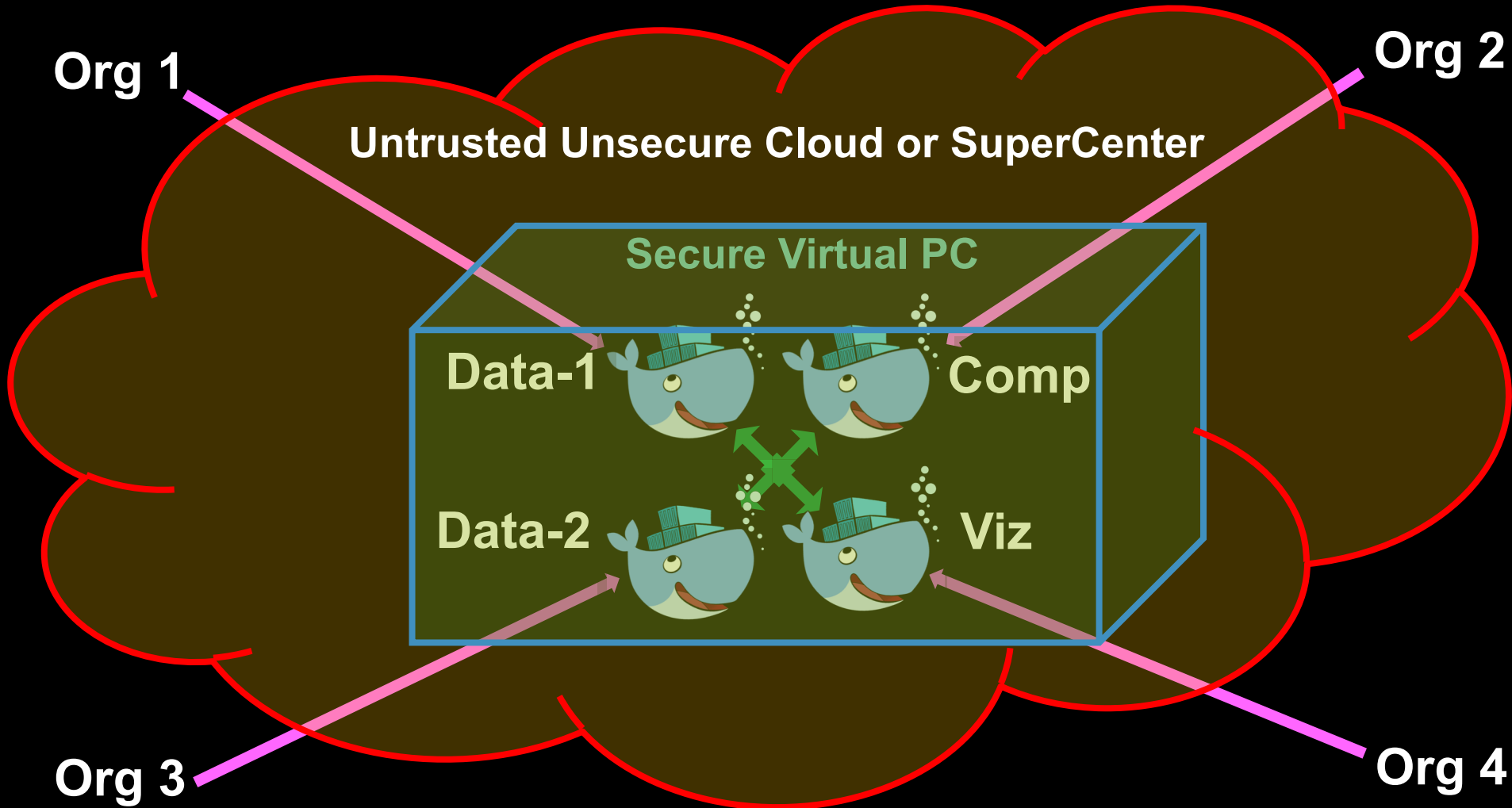
Layer 2 exchange service
Ethernet frames

ETH / ST

Secure Policy Enforced Data Processing



- Bringing data and processing software from competing organisations together for common goal
- Docker with encryption, policy engine, certs/keys, blockchain and secure networking
- Data Docker (virtual encrypted hard drive)
- Compute Docker (protected application, signed algorithms)
- Visualization Docker (to visualize output)



SC16 Demo

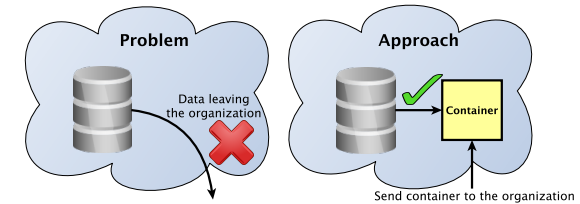
DockerMon Sending docker containers with search algorithms to databases all over the world.

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

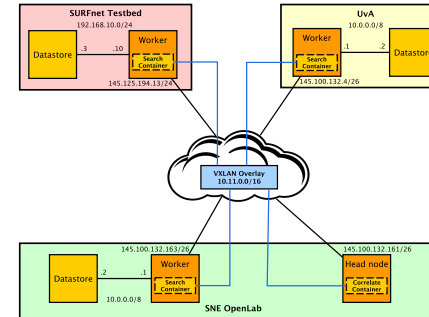
Container-based remote data processing

Problem Description

- Scientific datasets are usually made publicly available
...but data cannot always leave the organization premises
- On-site data processing can be challenging because of incompatibility of systems or lack of manpower
- Can a container-based system perform remote on-site data processing efficiently?
- What are the networking issues to solve?



Underlay and Overlay



Main features:

- Networked containers
- VXLAN overlay
- Containers that perform data retrieval and computation
- Containers built on-demand
- On-site data processing
- Distributed data source
- Multiple sites with datasets

The Game

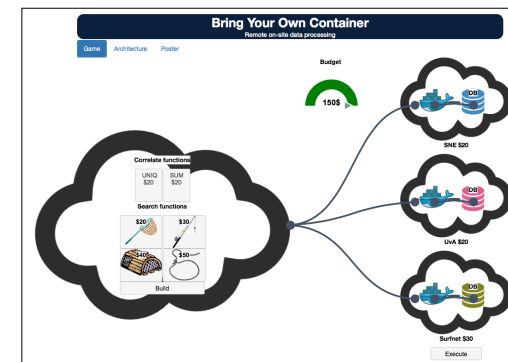
Our SC16 demo is a gamification of the remote dataset processing architecture.

How many different animal species can you find? You have a fixed budget and each function and processing will cost you money!

In our game you will:

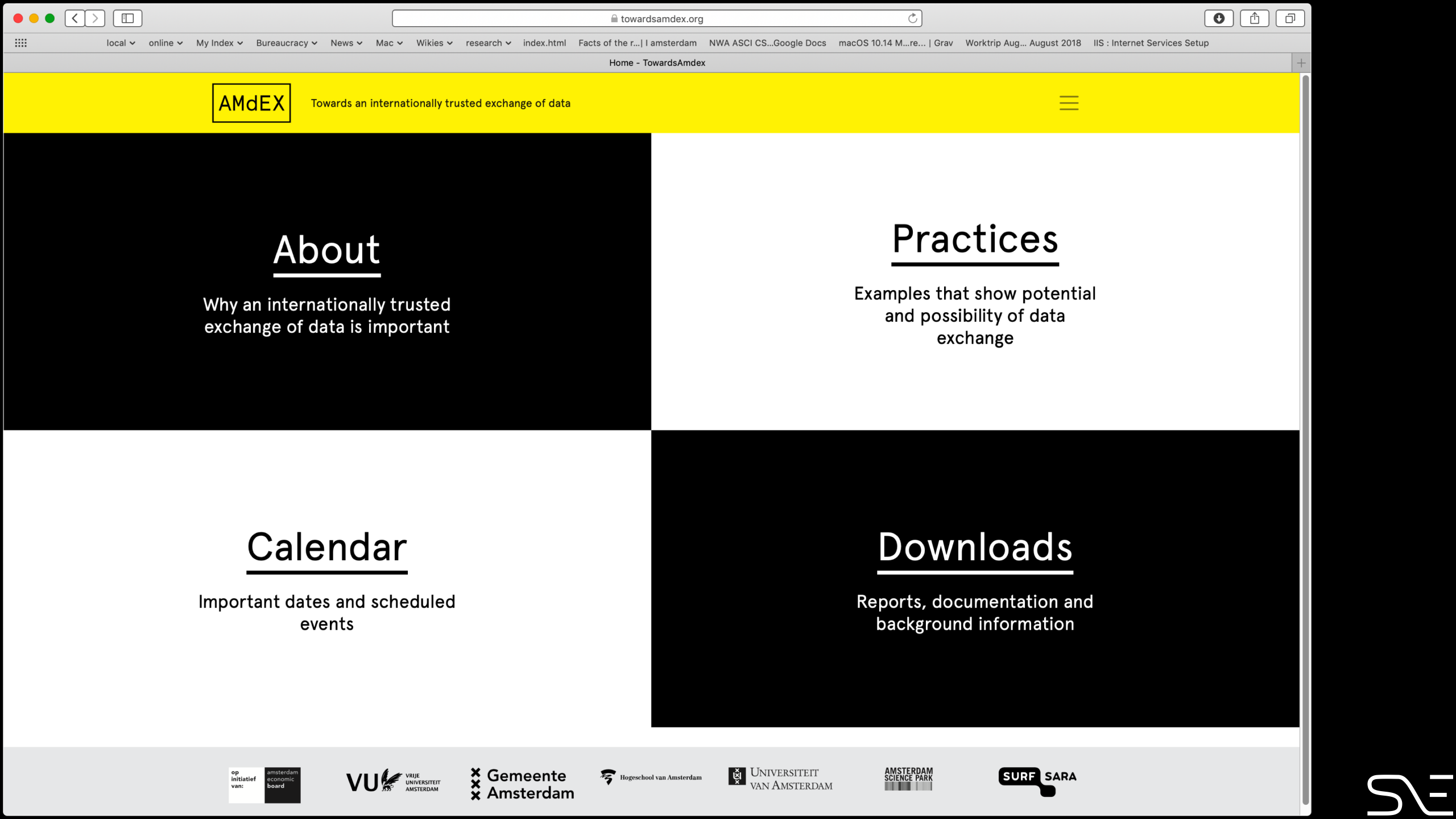
- Select a correlate function to combine the results of the different sites.
- Pick different search functions, represented as tools, to find animals in the remote datasets.
- Build containers with the search and correlate functions.
- Execute the containers on the sites of your choice.

Will you have the best score?



More information:

- <http://byoc.lab.uvalight.net/info>
- <http://sne.science.uva.nl/sne/gigaport3>
- <http://delaat.net/sc>



Towards an internationally trusted exchange of data



About

Why an internationally trusted exchange of data is important

Practices

Examples that show potential and possibility of data exchange

Calendar

Important dates and scheduled events

Downloads

Reports, documentation and background information



Q&A

- More information:
 - <http://delaat.net/dl4ld> and <http://delaat.net/epi>
 - <https://towardsamdex.org>
- Contributions from:
 - Leon Gommans, Wouter Los, Paola Grosso, Yuri Demchenko, Lydia Meijer, Tom van Engers, Reggie Cushing, Ameneh Deljoo, Sara Shakeri, Lu Zhang, Joseph Hill, Lukasz Makowski, Ralph Koning, Gleb Polevoy, Tim van Zalingen, and many others!

