Distributed FAIR information systems to enable federated learning and reasoning.

Applicants:
Cees de Laat (UvA), Henri Bal (VU), Barend Mons (LUMC, GO FAIR), Wessel Kraaij (UL/TNO Use Case)

Team:
Stef van Buuren (UU/TNO & RECAP), Paola Grosso (UvA), Jacopo Urbani (VU), Aske Plaat (UL), Tom van Engers (UvA), Sander Klous (KPMG/UvA), Leon Gommans (Air France KLM/UvA), Gert Kruithof (ASTRON), Peter Dieleman (NLR)
Collaboration Leiden - Amsterdam

Main RQ: How to construct a FAIR data service that allow competing organizations to share & analyse data for a common agreed goal but not for other purposes? (FAIR = Findable Accessible Interoperable Reusable)

UL / TNO Healthcare Use Case
LUMC – GO FAIR

Data Exchange

VU – Distributed Learning
UvA – Secure Data MarketPlace

Proof Of Concept on National System of Big Data Hubs
Inference and Reasoning

• It is crucial that next-gen information systems can extract new knowledge from the data

• Two paradigms for knowledge extraction
  • Reasoning -- leverage logic for providing output that is explainable and verifiable
  • Inference – leverage statistics for providing output that is robust to error and uncertainty

• Goal: Enable services for reasoning and inference on secure data-hubs and FAIR data

• To achieve this goal, we intend to exploit expertise on
  • Reasoning and inference on Semantically annotated data (VU)
  • Large-scale infrastructure and HPC (UvA & VU)
  • Realistic use-cases (health UL/TNO, Astronomy ASTRON, etc.)
Secure Trustworthy Digital Marketplaces (STDMPs)

- National Law & Regulations
- Market rules
- Adjudication
- Secure Digital Marketplace Member Organisation
- Agreement
- Registry
- Deployment Models
- Parameterization & authorizations
- Dispute Resolution
- Future Internet Capabilities
- Algorithm supplier(s)
- Data supplier(s)
- Customer(s)
- Accounting & Auditing
Use Case Personal Health Train

Thyroid Gland <> COPD?
- doctors: no
- Pubmed: no

Thyroid Gland <> COPD?
- EURETOS: YES
- measure!

Thyroid Gland <> COPD?
- 196.425 K patients
- predisposition? CLCX13

Thyroid Gland <> COPD?
- measure!

Hospital 1
- Measure CLCX13

Hospital 2
- Measure CLCX13

Hospital n
- Measure CLCX13
Data Processing models

- Bring data to computing
- Bring computing to data
- Bring computing and data to (un)trusted third party
- A mix of all of the above
- Block chain to record what happened
- Block chain for data integrity
- Bring the owner of Data in control!
- Data owner policy + enforcement technology
The VMs that are live-migrated run an iterative search-refine-search workflow against data stored in different databases at the various locations. A user in San Diego gets hitless rendering of search progress as VMs spin around.

Validation Fieldlab and Dissemination

- Experimental facilities from day one!
- Proof of concepts demonstrating secure data sharing
- Blueprint, roadmap and standards where applicable
- Model for FAIR EOSC Infrastructure

UVA - OpenLab
- KLM
- NetherLight
- GENI
- Fed4Fire
- Cloud
- SURFSARA
- ...

TNO - Intrepid
- Smart Data Factory
- Innovations
- Smart Rail
- To-Grip
- ...

C2D – Big DataHubs
- Arena
- KAVE
- AZURE
- Use Cases
- ...

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

PRP Partners include:
- Univ. of Hawaii System
- Montana State Univ.
- Northwestern Univ.
- NCAR
- MREN
- StarLight
- UIC
- Chameleon
- UvA
- AARNet
- KISTI/KREONet
- Univ. of Tokyo
- NCSA
- Clemson Univ.

Note: this diagram represents a subset of sites and connections.
Main deliverables

- Generic federated analytics using secure data hub concepts applied to health use case.
- Proof of principle for distributed inference at FAIR secure data hubs.
- Validation at a secure data-hub science venue and validation at life science venue with data from repositories in three countries.

Proof Of Concept on National System of Big Data Hubs
Contribution to the program.

Keywords
• distributed learning, secure workflows, standard setting, international embedding.....

Potential collaborations
• WP3 (Jacobs et al) = PEP and datasets that need to ‘travel the circuit’
• WP8: Dumontier et al.
• All three projects work with secure and privacy sensitive data and how to learn from them with respect for integrity.
• P7 and P8 have already indicated to form a joint development team is possible

Overarching FACT/FAIR and Big Data at large
• In a certain sense our project(s) address the overlapping aspects of FACT and FAIR. The developments will be guided by FAIR principles, which will ensure that aspects of FACT can be implemented technically.

Funding opportunities
• We are developing both national (LSH) PPP’s and a large international consortium, (BCG/WEF, Germany, Switzerland) to implement a reference implementation of the FAIR/PHT broad principles for value based health care and beyond.
• Smart Industry & City & EOSC (DL4LD, SARNET)