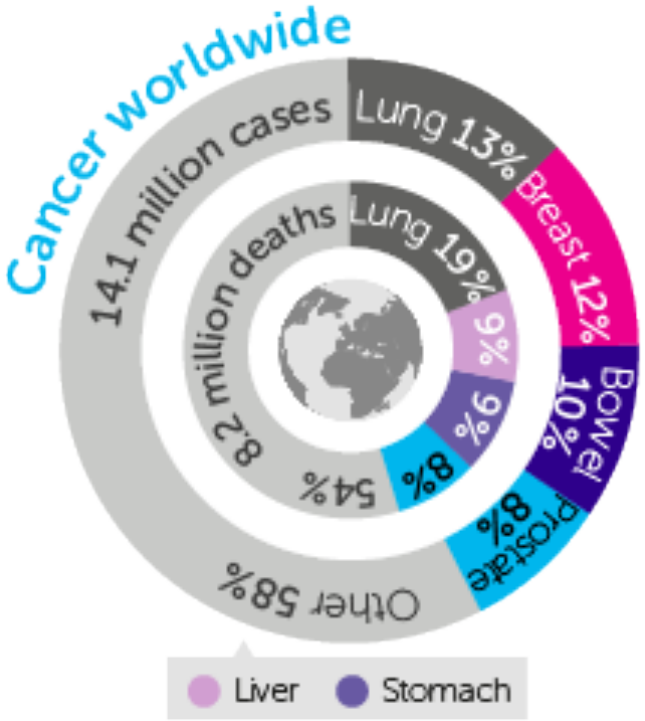

PERSONAL
HEALTH TRAIN
The Personal Health Train
Privacy Preserving Federated Data Analysis
Ananya Choudhury
Maastro Clinic, Maastricht
Maastricht University, Netherlands
Maastro
Maastricht University
Maastricht UMC+

Big Data – An example for Cancer



Oncology
2007-2017
150M patients
0.1-10GB per patient

15-1500PB
80% unstructured

Hospitals
China: 25.000
India: 35.000
Germany: 2.000
France: 2.300
Italy: 1.100
USA: 5.500
Australia: 1.400

TOTAL ~100.000

The Health Data GoldiLocks Dilemma

SHARING ??? **PRIVACY ???**

BOTH ????

- Broader Data Interoperability and data sharing

  and / or

- Enhanced Data Privacy

Deven McGraw & Vince Kuraitis,
Health 2.0, September 18, 2019, California , The USA.

Barriers to sharing data

[..] the problem is not really technical […]. Rather, the problems are **ethical, political, and administrative**.

Lancet Oncol 2011;12:933

1. Administrative (I don't have the resources)
2. Political (I don't want to)
3. Ethical (I am not allowed to)

4. Technical (I don't know how)

Solution to the Dilemma:

DO NOT share data

Instead, send **applications** and **results**

Personal Health Train – Infrastructure to send applications and results

Challenges

- The research application has to be distributed (trains & track)
- The data has to be understandable by an application (i.e. not a human) -> FAIR data stations

Example:

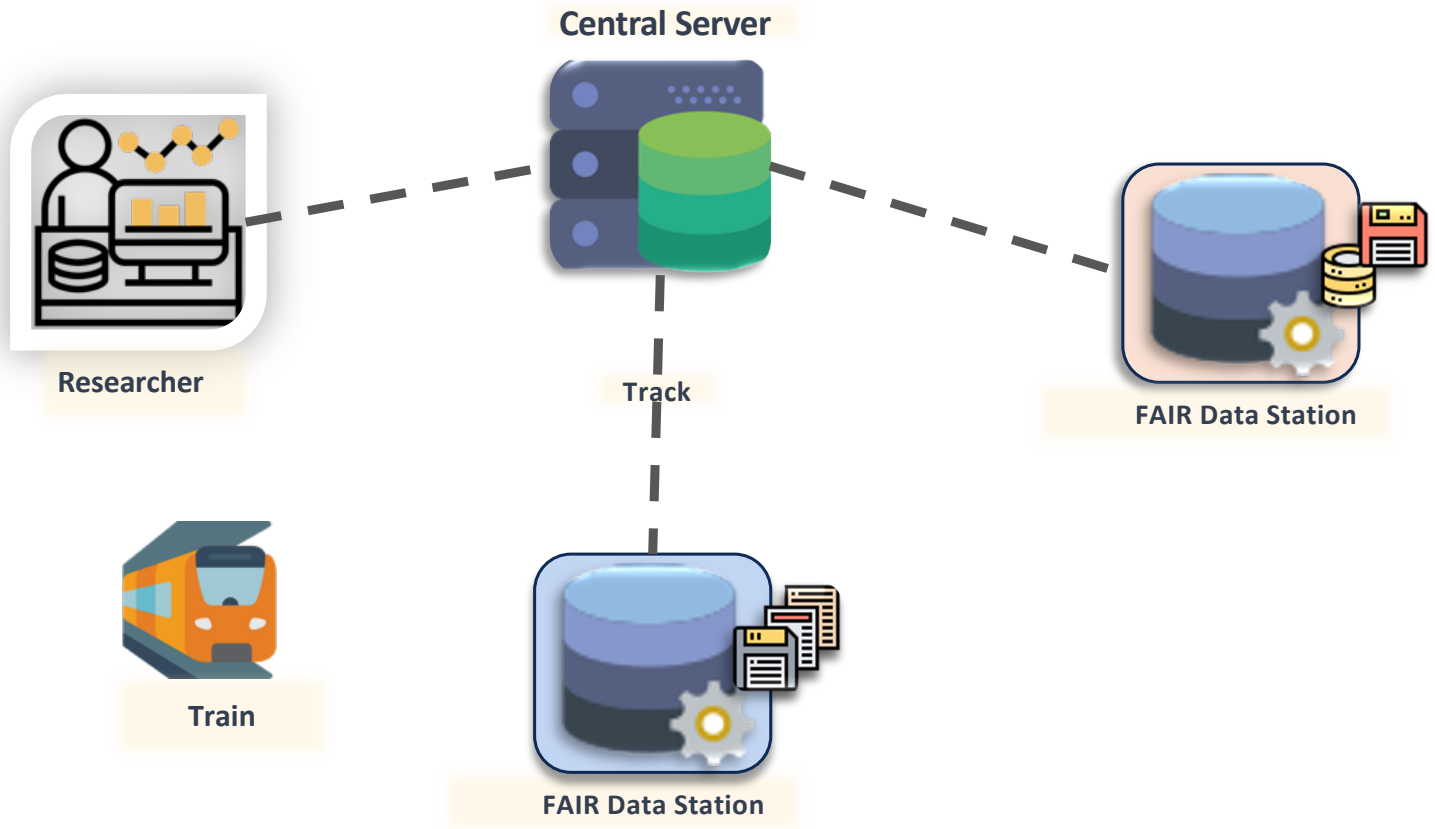
who3g	bmi	fev1pc_t0	dumsmok	t_ct_loc	tstage	nstage	hist4g
2	23	82	2	6	4	4	4
2	29	62	2	1	4	3	1
1	25	95	2	4	4	1	1
2	29	73	2	6	0	4	2
1	20	124	2	6	2	3	4
2	22	56.8	2	1	0	3	4
2	23	76	2	6	2	3	3

IsSCLC	T_stage	N_stage	M_stage	PA	Locatie	FEV	Cur
0	0	2	0	0	1	97	
0	3	2	0	1	0	61	
0	1	3	0	1	0		
0	1	3	0	1	0	91	
0	1	2	0	1	2		
0	3	2	0	1	2	89	

https://www.cancerdata.org

Personal Health Train:

https://distributedlearning.ai/blog/



Components of PHT

Tracks

Routes for application/results transportation

Station

Trains enter station via tracks

- Execution environment for train
- FAIR Data !?

Train

Application / Algorithm / Analysis Scripts

Logic of the application
- Data Query
- Analysis executed on the queried data

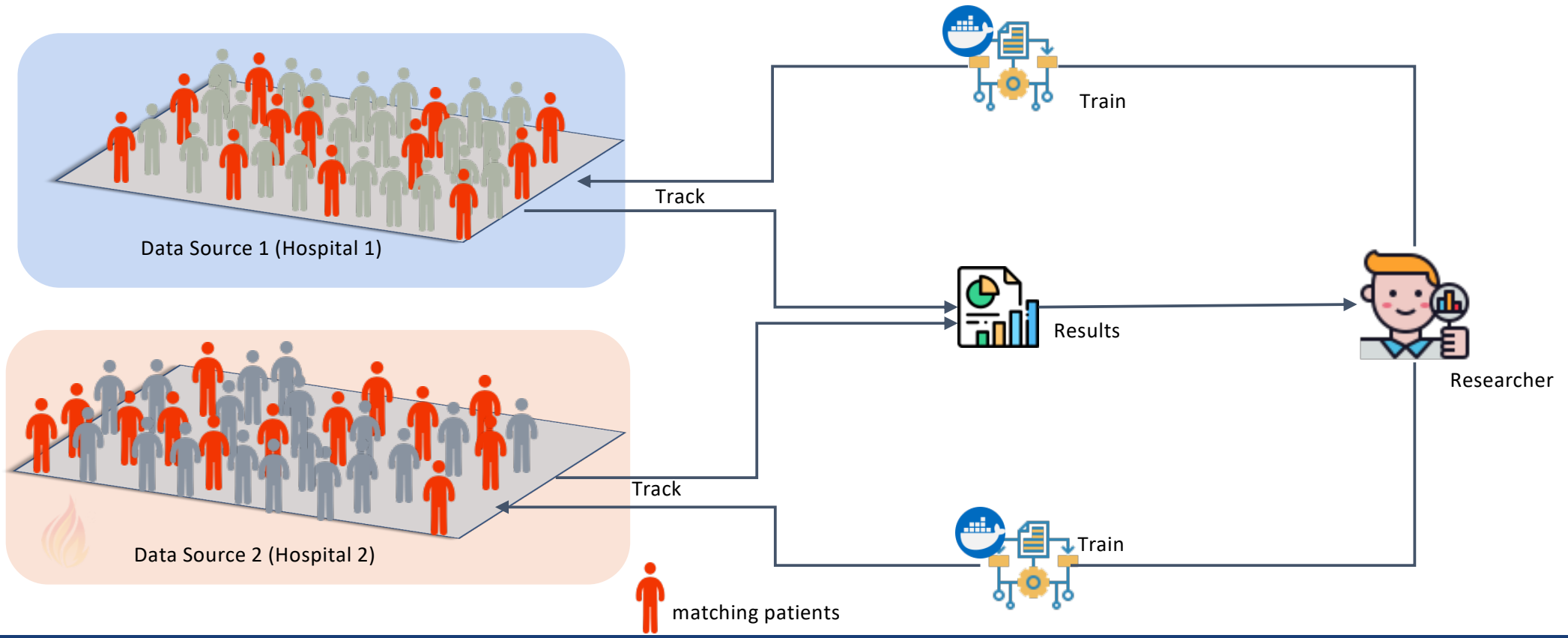
Animation

https://vimeo.com/143245835

https://vimeo.com/143245835

Use Cases and Applications

Patient Cohort Analysis from Multiple Data Sources



Applications

- Distributed Learning on **20,000+ lung cancer patients** – *Radiotherapy and Oncology Journal , 2020*
  - 2 year survival prediction
  - 5 countries, 8 healthcare institutes – Amsterdam, Cardiff, Maastricht, Manchester, Nijmegen, Rome, Rotterdam, Shanghai

- Calculating Healthcare Quality Indicators from distributed datasets for colorectal patients

- Survival Prediction for anal cancers from distributed datasets - Maastricht, Oslo, Leeds

- **Distributed Deep Learning** with PHT for survival prediction from CT images

Publications

Original Article

Distributed learning on 20 000+ lung cancer patients – The Personal Health Train

Timo M. Deist [a,b,1], Frank J.W.M. Dankers [a,c,1], Priyanka Ojha [d], M. Scott Marshall [d], Tomas Janssen [d], Corinne Faivre-Finn [e], Carlotta Masciocchi [f,g], Vincenzo Valentini [f,g], Jiazhou Wang [h], Jiayan Chen [h], Zhen Zhang [h], Emiliano Spezi [i,j], Mick Button [j], Joost Jan Nuyttens [k], René Vernhout [k], Johan van Soest [a], Arthur Jochems [b], René Monshouwer [c], Johan Bussink [c], Gareth Price [e,2], Philippe Lambin [b,2], Andre Dekker [a,2,*]

[a] Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+; [b] The D-Lab: Dpt of Precision Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+; [c] Department of Radiation Oncology, Radboud University Medical Center, Nijmegen; [d] Department of Radiation Oncology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek, Amsterdam, The Netherlands; [e] The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, United Kingdom; [f] Università Cattolica del Sacro Cuore; [g] Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy; [h] Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; [i] School of Engineering, Cardiff University; [j] Velindre Cancer Centre, Cardiff, United Kingdom; [k] Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

SCIENTIFIC DATA

OPEN ARTICLE

Distributed radiomics as a signature validation study using the Personal Health Train infrastructure

Zhenwei Shi [1,7*], Ivan Zhovannik [1,2,7], Alberto Traverso [1,6], Frank J. W. M. Dankers [1,2], Timo M. Deist [1,3], Petros Kalendralis [1], René Monshouwer [2], Johan Bussink [2], Rianne Fijten [1], Hugo J. W. L. Aerts [4,5], Andre Dekker [1] & Leonard Wee [1]

Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data

Johan van SOEST[a,1], Chang SUN[b], Ole MUSSMANN[c], Marco PUTS[c], Bob van den BERG[c], Alexander MALIC[b], Claudia van OPPEN[b], David TOWEND[d], Andre DEKKER[a] and Michel DUMONTIER[b]

DATA Intelligence

Quarterly
Founded: 2018
E-ISSN: 2641-435X

Distributed Analytics on Sensitive Medical Data: The Personal Health Train

Show all authors
Oya Beyan, Ananya Choudhury, Johan van Soest, Oliver Kohlbacher,

Posted Online January 31, 2020

Other Areas

- Different Types of Cancer
- Alzheimer’s / Dementia
- COVID-19
- Cardiovascular Disease Prevention


What is the PHT?
health RI
enabling data driven health
ata Driven Health
Connect, Share and Reuse


HEALTH-RI CONFERENCE
Towards data driven health
January 30th, 2020 | Jaarbeurs Supernova Utrecht

Use cases

www.personalhealthtrain.nl

Acknowledgement

Netherlands
- MAASTRO, Maastricht, Netherlands
- Radboudumc, Nijmegen, Netherlands
- Erasmus MC, Rotterdam, Netherlands
- Leiden UMC, Leiden, Netherlands
- Catharina Hospital, Eindhoven, Netherlands
- Isala Hospital, Zwolle, Netherlands
- NKI Amsterdam, Netherlands
- UMCG, Groningen, Netherlands
- IKNL, Utrecht, Netherlands

Europe
- Policlinico Gemelli & UCSC, Roma, Italy
- UH Ghent, Belgium
- UZ Leuven, Belgium
- Cardiff University & Velindre CC, Cardiff, UK
- CHU Liege, Belgium
- Uniklinikum Aachen, Germany
- LOC Genk/Hasselt, Belgium
- The Christie, Manchester, UK
- State Hospital, Rovigo, Italy
- St James Institute of Oncology, Leeds, UK
- U of Southern Denmark, Odense, Denmark
- Greater Poland Cancer Center, Poznan, Poland
- Oslo University Hospital, Oslo, Norway

Africa
- University of the Free State, Bloemfontein, South Africa

Asia
- Fudan Cancer Center, Shanghai, China
- CDAC, Pune, India
- Tata Memorial, Mumbai, India
- HGC Oncology, Bangalore, India

North America
- RTOG, Philadelphia, PA, USA
- MGH, Boston, MA, USA
- University of Michigan, Ann Arbor, USA
- Princess Margaret CC, Canada

South America
- Albert Einstein, Sao Paulo, Brazil

Australia
- University of Sydney, Australia
- Westmead Hospital, Sydney, Australia
- Liverpool and Macarthur CC, Australia
- ICCC, Wollongong Australia
- Calvary Mater, Newcastle, Australia
- North Coast Cancer Institute, Coffs Harbour, Australia

Industry
- Varian, Palo Alto, CA, USA
- Philips, Bangalore, India
- Sohard GmbH, Fuerth, Germany
- Microsoft, Hyderabad, India
- Mirada Medical, Oxford, UK
- CZ Health Insurance, Tilburg, NL
- Siemens, Malvern, PA, USA
- Roche, Woerden, NL
- Medical Data Works, Heerlen, NL