

The Personal Health Train

Privacy Preserving Federated Data Analysis

Ananya Choudhury

Maastro Clinic, Maastricht

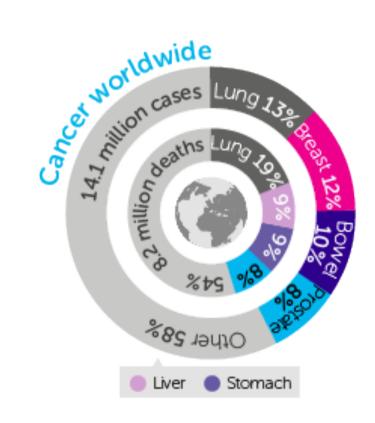
Maastricht University, Netherlands







Big Data – An example for Cancer



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

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:

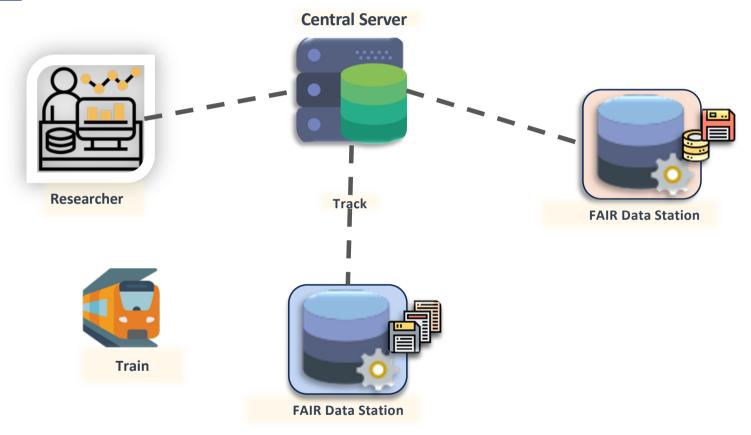
who3g	bmi	fev1pc_t0	dumsmok	t_ct_loc	tstage	nstage	hist4g
2	23	82	2	6	4	4	4
2	29	62	2	1	4	3	1
1	25	95	2	4	4	1	1
2	29	73	2	6	0	4	2
1	20	124	2	6	2	3	4
2	22	56.8	2	1	0	3	4
2	23	76	2	6	2	3	3
	-		132		192	92	100

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	
_	_	_	_		_		



Personal Health Train:

https://distributedlearning.ai/blog/





Components of PHT



Tracks
Outes for application

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

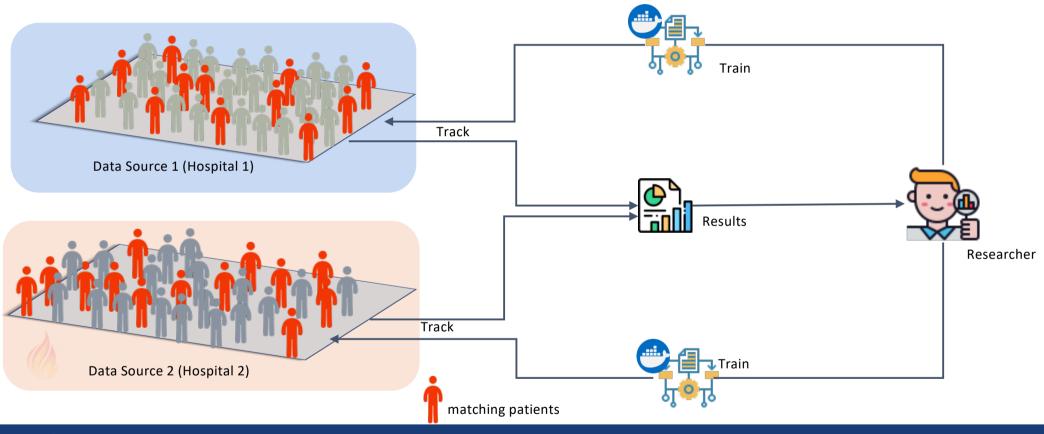
https://vimeo.com/143245835



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^g, Vincenzo Valentini^{f,g}, Jiazhou Wang^h, Jiayan Chen^h, Zhen Zhang h, Emiliano Spezi i J, Mick Button J, Joost Ian Nuvttens 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,*

*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+; Department of Radiation Oncology, Radboud University Medical Center, Nijmegen; ^d Department of Radiation Oncology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek, Amsterdam, The Netherlands; ^eThe University of Manchester, Manchester Academic Health Science Centre. The Christie NHS Foundation Trust, United Kingdom: Università Cattolica del Sacro Cuore: 8 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; School of Engineering, Cardiff University; Welindre Cancer Centre, Cardiff, United Kingdom; Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam. The Netherlands

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 DEKKERa and Michel DUMONTIER b

SCIENTIFIC DATA

OPEN Distributed radiomics as a signature **ARTICLE** 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 Kalendralis¹, René Monshouwer², Johan Bussink², Rianne Fijten¹, Hugo J. W. L. Aerts4,5, Andre Dekker1 & Leonard Wee1

Home | Data Intelligence | List of Issues | Volume 2 , No. 1-2 | Distributed Analytics on Sensitive Medical Data: The Personal Health Train



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

Distributed Analytics on Sensitive Medical Data: The Personal Health Train

Oya Beyan O, 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









Use cases www.personalhealthtrain.nl

Health care

Simplifying the Risk

A possible future project that aims to replace the current system of collecting data for risk equalisation. Using artificial into support treates
s decisions in me

e aim of this possible future oject is to provide objective toome measures that enable determine the best OzCAT: The Australian Computer Aided Theragnostics network

The aim of this data-network is to enable the generation of additional clinical evidence to support treatment decisions for patients by machine learning of

of Limburg Meet (LIME)

The aim of this project is to test.

AMICUS: AI in Medical Imaging for Novel Cance User Support

Develop and use technology for privacy-preserving distributed deep learning from existing nospital imaging archives. A Survival Model for NSCLC nrough I Learning Across

The aim of this project is to sevelop a predictive model of survival at 2 years based on a arge volume of historical outlent data that serves as a proof of concept to demonstrate the distributed learning approach.

he Dutch Network of Computer Assisted Theragnostics (duCAT)

Develop and validate a framework for clinical cancer research, incorporating predictive models developed by machine learning based on data from multiple institutions. euroCAT: Distributed Learning for Individualize Medicine

nfrastructure and distributed earning methodology for rivacy-preserving multi-centric apid learning health care thruption nchmarking in health e by secure Multi Pa mputation

pare calculations on I quality by secure rty computation with to a traditional quality 20k Challenge - A challenge to use the data of 20000 lung cancer patients from 5 countries to predict outcomes of therapy

The aim is to show that the PHT distributed learning infrastructure can be scaled to many thousands of patients from at least five healthcare providers in more than five PROSPECT: personalize health decision aid facilitated by the PHT

Development of a prostate cancer decision aid, with personalized content facilitated chinaCAT: Rapid learning infrastructure for outcome prediction models in rectal

Developing a rapid learning infrastructure for rectal cancer

My Best Treatment: datadriven decision support with terminal lung cancer

> The aim of the project is to set up an infrastructure for automatic data registration related to novel proton theraps in the Metharlands

CONVINCED: analysis of vertically partitioned data using multi-party computation

Enabling survival analysis on vertically partitioned data wh securing privacy using multiparty computation. Infrastructure Rare Cancers in E nal Health Asia (RARECARE

ie aim of this project is to welop and deploy rails for the rrsonal Health Train. Vantage6 an open source infrastructure connect data stations and run Understanding Oral Cavit cancer survival in the Netherlands and Taiwan

Federated analysis of survipatients with oral cavity can based on cancer registry da

Decentralised International Cancer Surveillance and Comparison using the Personal Health Train

The aim of this project is to provide a privac preserving solution for international cancer statistics and surveillance programs. rediction model for prostate cancer

The aim of this project is to create a prediction model for survival from prostate cancer, based on cancer registry data from at least two geographies

← Use cases Healthy living

ersonal Health Train, pa of Limburg Meet (LIME)

The aim of this project is to te the Personal Health Train in practice. NVINCED: analysis of tically partitioned data ng multi-party nputation

bling survival analysis on scally partitioned data while uring privacy using multity computation rope and Decentralised Internat et Asia) Cancer Surveillance an Comparison using the

> The aim of this project is to provide a privacy-preserving solution for international cancer statistics and surveillance programs.

Health research

namic consent in a

Pe of

The aim of this project is to demonstrate how individual citizens can provide or deny access to their personal data AMICUS: Al i

n of this project is to test rsonal Health Train in se. : Al in Medical meerCAT: A for Novel Cancer Prediction port Patients Th

> The aim of this project is to develop a predictive model of survival at 2 years based on a large volume of historical patient data that serves as a proof of concept to demonstrate the distributed learning approach.

The Dutch Network of Computer Assisted Theragnostics (duCAT)

Develop and validate a framework for clinical cancer research, incorporating predictive models developed by machine learning based on data ing for Individualized Benchmarki

To compare calculations of hospital quality by secure multiparty computation w respect to a traditional qui benchmark calculation

20k Challenge - A challeng to use the data of 20000 lung cancer patients from a countries to predict outcomes of therapy

a with The aim is to show that the quality distributed learning intrastructure can be scale many thousands of patien from at least five healthca providers in more than five

infrastructure for outcome prediction models in rectal cancer

eveloping a rapid learning frastructure for rectal cancer with terminal lung cancer

Data-driven decision support fo

n decision support for The aim of the projecth terminal lung up an infrastructure automatic data regis related to novel prot

CONVINCED: analysis of vertically partitioned data using multi-party

et

Enabling survival analysis vertically partitioned data securing privacy using mul

Asia (RARECAREnet A

AREnet Asia) cancer survival Netherlands ar

Federated analysis of survival of patients with oral cavity cancer based on cancer registry data. Decentralised International Cancer Surveillance and Comparison using the

> he aim of this project is to rovide a privacy-preserving slution for international cancer tatistics and surveillance rograms.



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
- State Hospital, Novigo, 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



