The Personal Health Train

Privacy Preserving Federated Data Analysis

Ananya Choudhury

Maastro Clinic, Maastricht

Maastricht University, Netherlands
Big Data – An example for Cancer

Cancer worldwide

141 million cases
82 million deaths
8% Liver
54% Stomach
20% Other
19% Lung
12% Breast
9% Brain
9% Prostate
8% Bowel

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:

<table>
<thead>
<tr>
<th>who3g</th>
<th>bmi</th>
<th>fev1pc_t0</th>
<th>dumsmok</th>
<th>t_ct_loc</th>
<th>tstage</th>
<th>nstage</th>
<th>hist4g</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23</td>
<td>82</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>62</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>95</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>73</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>124</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>56.8</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>76</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IsSCLC</th>
<th>T_stage</th>
<th>N_stage</th>
<th>M_stage</th>
<th>PA</th>
<th>Locatie</th>
<th>FEV</th>
<th>Cur</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>89</td>
<td></td>
</tr>
</tbody>
</table>

[https://www.cancerdata.org](https://www.cancerdata.org)
Personal Health Train:
https://distributedlearning.ai/blog/
Components of PHT

Tracks
Routes for application/results transportation

Train

Station

Trains enter station via tracks
- Execution environment for train
- FAIR Data !?

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

Data Source 1 (Hospital 1)

Data Source 2 (Hospital 2)

matching patients

Train

Track

Results

Researcher
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


*Department of Radiation Oncology (MASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre. 
†The Dutch E-line of Decision Making, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre. 
‡Department of Radiation Oncology, Radboud University Medical Center, Nijmegen. 
§Department of Radiation Oncology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek, Amsterdam, The Netherlands. 
¶The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, United Kingdom. 
©Universitaria Clinica del Sasso Cancer Center, Fondazione Policlinico Universitaria A. Gemelli IRCCS, Rome, Italy. 
©Department of Radiation Oncology, Peking University Thoracic Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China. 
*School of Engineering, Cardiff University. 
††Vinderup Cancer Centre, Cardiff, United Kingdom. 
§§Department of Radiation Oncology, University of Southern California, Los Angeles, California, USA.

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

Johan van SOEST, Chang SUN, Ole MUSSMANN, Marco PUTS, Bob van den BERG, Alexander MALIC, Claudia van OPPEN, David TOWEND, Andre DEKKER and Michel DUMONTIER.

Distributed Analytics on Sensitive Medical Data: The Personal Health Train

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

Published Online January 31, 2020
Other Areas

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

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