

# Advances in clinical informatics for outcome research Federated learning

Andre Dekker

Medical Physicist | Professor of Clinical Data Science  
Maastricht UMC+ | Maastricht University | Maastricht Clinic

Advances in clinical informatics for outcome research  
DIGICORE Connect To Win 2022 | Milan | Nov 9, 2022 | 10:30-11:00

# Disclosures

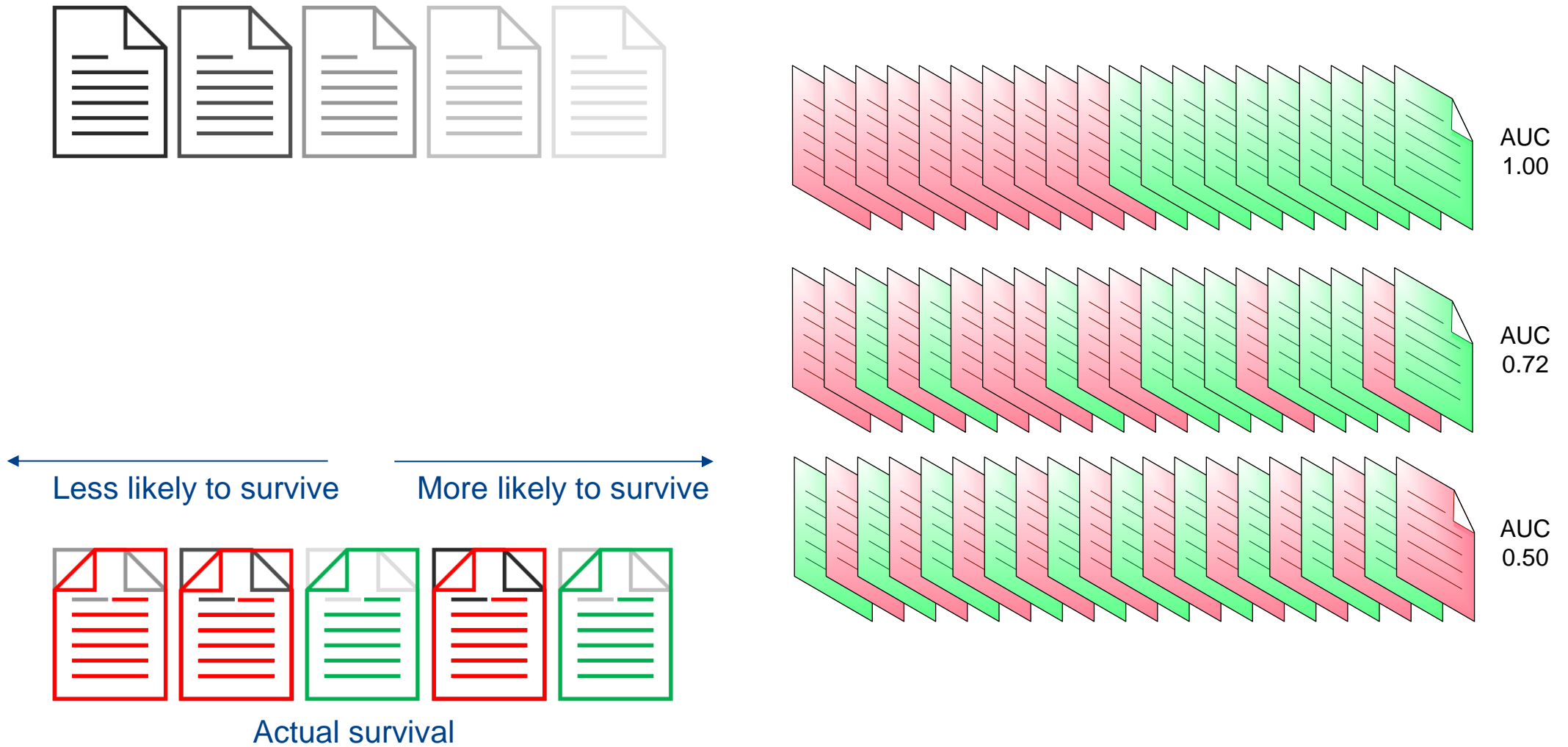
Research collaborations incl. funding, consultancy and speaker honoraria

- Pharma: Roche, Janssen, Bristol-Myers Squibb
- MedTech/Data: Varian Medical Systems, Siemens, Philips, Sohard, Mirada Medical, ptTheragnostics, OncoRadiomics, IQVIA
- Health insurance: CZ Health Insurance

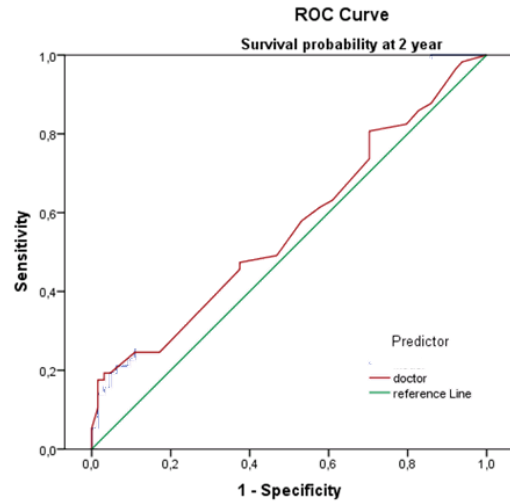
Spin-offs and commercial ventures

- MAASTRO Innovations B.V.
- Medical Data Works B.V.
- Various patents on medical machine learning & Radiomics

# Prediction of survival

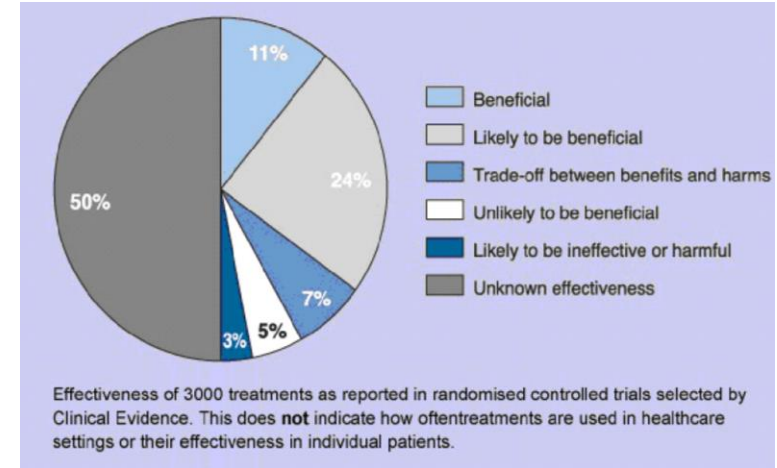
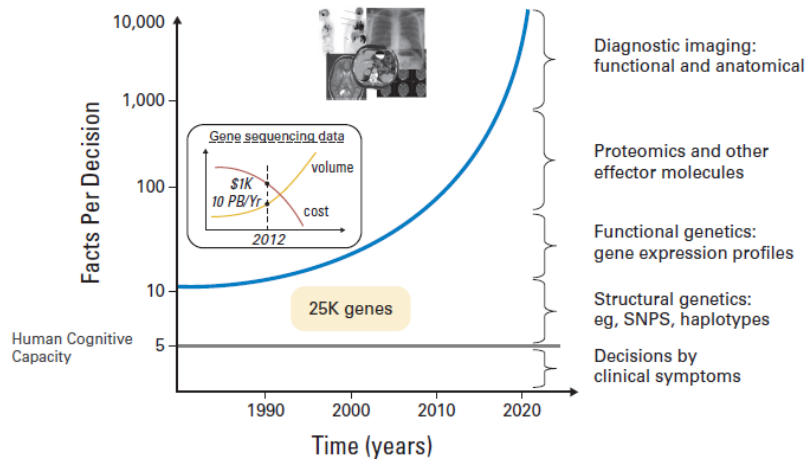


# Prediction of individual outcomes – we are drowning



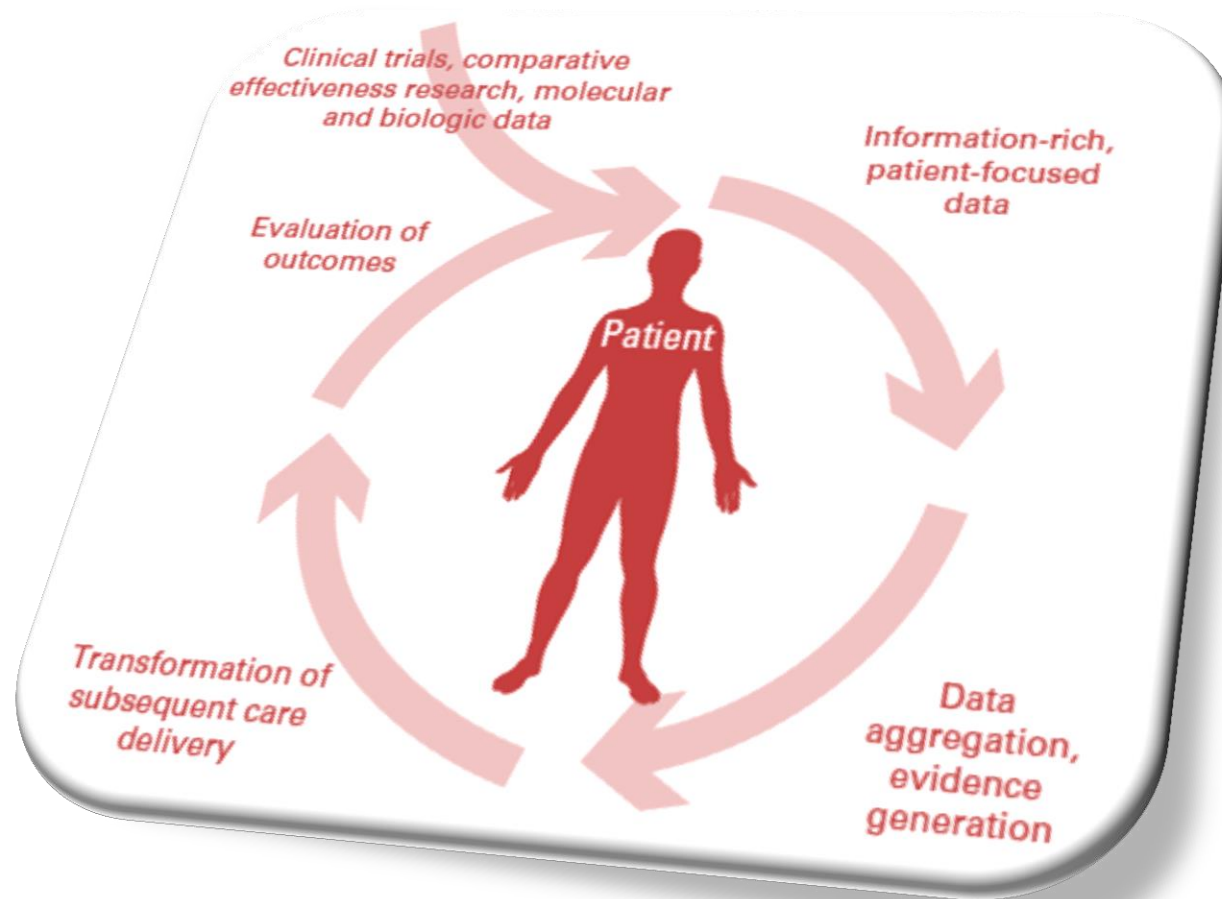
NSCLC (Lung Cancer)  
2 year survival  
158 patients  
5 MDs  
Prospective  
**AUC: 0.56**

- Explosion of data
- Explosion of decisions
- Explosion of 'evidence'
  - Too much to read
  - 3 % in trials, bias
  - Sharp knife



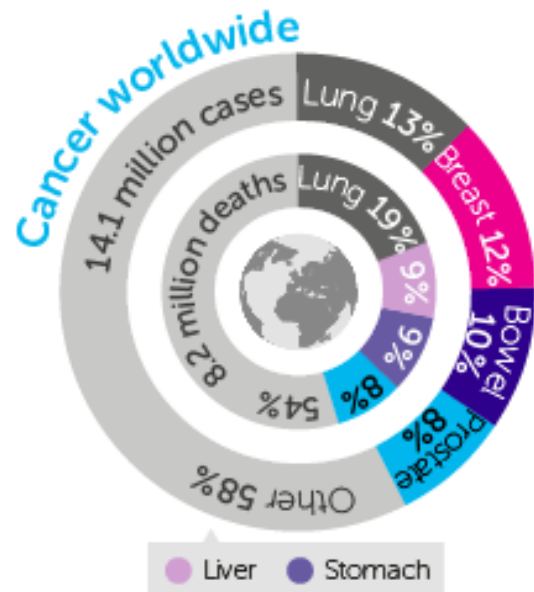
Oberije et al. , Radiother Oncol. 2014; 112: 37–43 / J Clin Oncol 2010;28:4268 / JMI 2012 Friedman, Rigby / BMJ Clinical Evidence

# Potential of Real World Data & Artificial Intelligence Learning Health Care System – Faster Innovations & Better Outcomes



## Data

# The main problem of health data is that it is fragmented



**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**

# 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 can't)

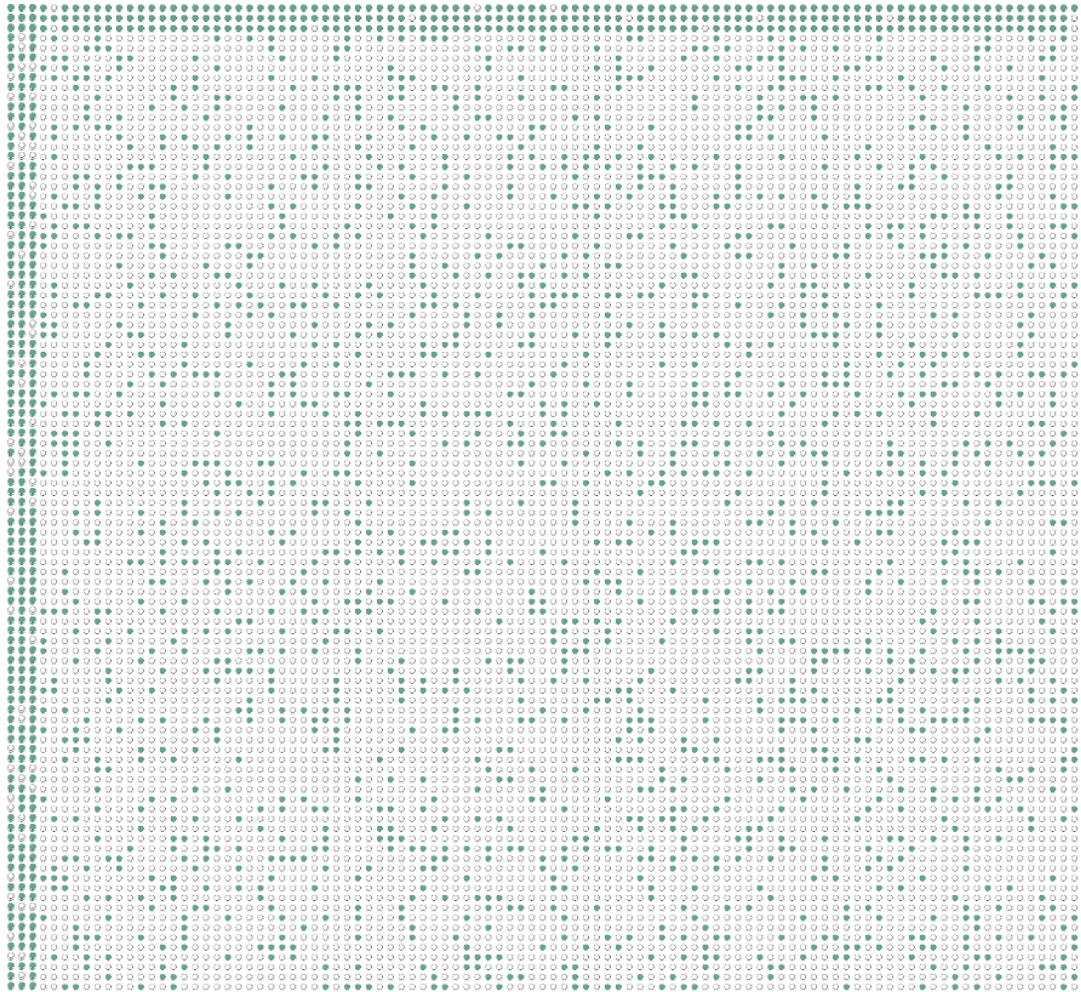




# Data landscape

## Data elements

Patients



- Clinical research
  - 3% of patients
  - 100% of features
  - 5% missing
  - 285 data points
- Clinical registries
  - 100% of patients
  - 3% of features
  - 20% missing
  - 240 data points
- Clinical routine
  - 100% of patients
  - 100% of features
  - 80% missing
  - 2000 data points

# A different approach

- If sharing is the problem: Don't share the data
- If you can't bring the data to the research
- You have to bring the research to the data
- Challenges
  - The research application has to be distributed (trains & track)
  - The data has to be understandable by an application (i.e. not a human) -> data stations

# Federated Learning – Maastricht Short History

*2007: The “Computer Assisted Theragnostics” (CAT) project will use heterogeneous data from distributed databases in multiple clinical centres to develop and validate patient specific prediction models”.*

*2008: First patent application (with Siemens) “System and method for privacy preserving predictive models for lung cancer survival analysis”  
“A framework [that] enables designing/learning improved predictive models that perform better than the individual models obtained by using local data from only one institution, while addressing the local and international privacy preserving concerns that arise when sharing patient related data.”*

*2014-2015: SWAT4LS Berlin & Leiden FAIR data*



- (22) **International Filing Date:** 16 January 2009 (16.01.2009)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**

61/022,009	18 January 2008 (18.01.2008)	US
12/353,310	14 January 2009 (14.01.2009)	US
- (71) **Applicants (for all designated States except US):**  
**SIEMENS MEDICAL SOLUTIONS USA, INC.**  
 [US/US]; 51 Valley Stream Parkway, Malvern, Pennsylvania 19355-1406 (US). **MAASTRO CLINIC** [NL/NL];  
 Dr. Tanslaan 12, NL-6229 ET Maastricht (NL).

# Personal Health Train (2015)

*Personal Health Train*

[Open Access](#) | [Published: 15 March 2016](#)

## **The FAIR Guiding Principles for scientific data management and stewardship**

[Mark D. Wilkinson](#), [Michel Dumontier](#), [IJsbrand Jan Aalbersberg](#), [Gabrielle Appleton](#), [Myles Axton](#), [Arie Baak](#), [Niklas Blomberg](#), [Jan-Willem Boiten](#), [Luiz Bonino da Silva Santos](#), [Philip E. Bourne](#), [Jildau Bouwman](#), [Anthony J. Brookes](#), [Tim Clark](#), [Mercè Crosas](#), [Ingrid Dillo](#), [Olivier Dumon](#), [Scott Edmunds](#), [Chris T. Evelo](#), [Richard Finkers](#), [Alejandra Gonzalez-Beltran](#), [Alasdair J.G. Gray](#), [Paul Groth](#), [Carole Goble](#), [Jeffrey S. Grethe](#),  
... [Barend Mons](#)  [+ Show authors](#)

[Scientific Data](#) **3**, Article number: 160018 (2016) | [Cite this article](#)

**500k** Accesses | **4834** Citations | **2037** Altmetric | [Metrics](#)

# Google (2017)

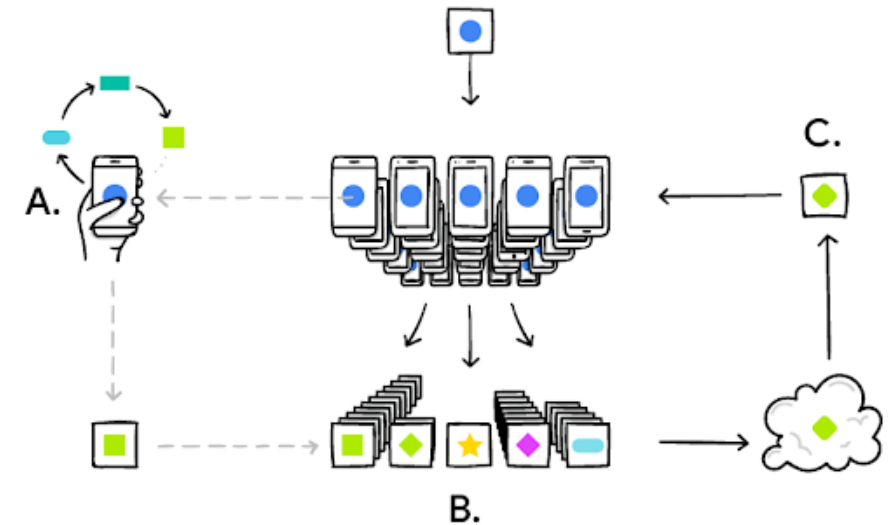
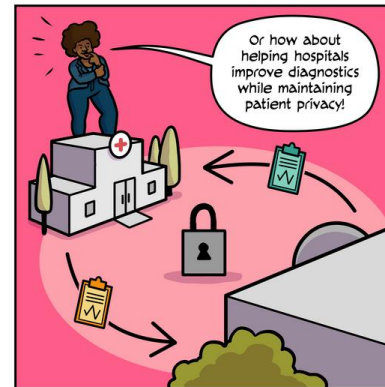
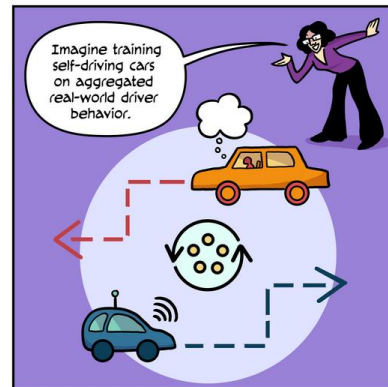
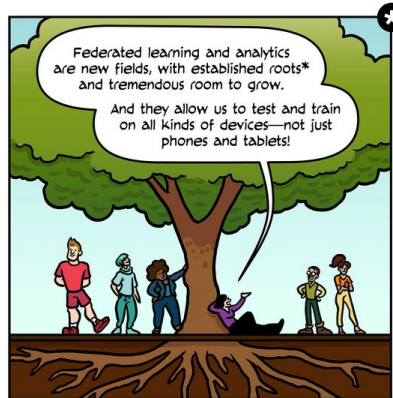
Google Research   Philosophy   Research Areas   Publications   People   Tools & Downloads

BLOG >

## Federated Learning: Collaborative Machine Learning without Centralized Training Data

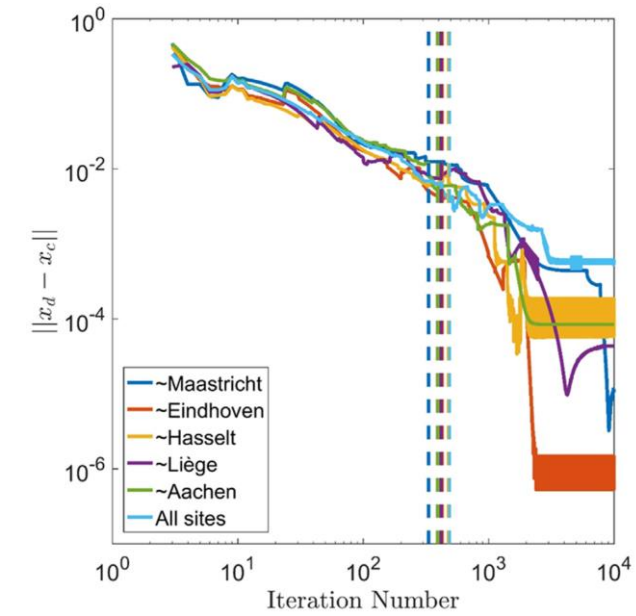
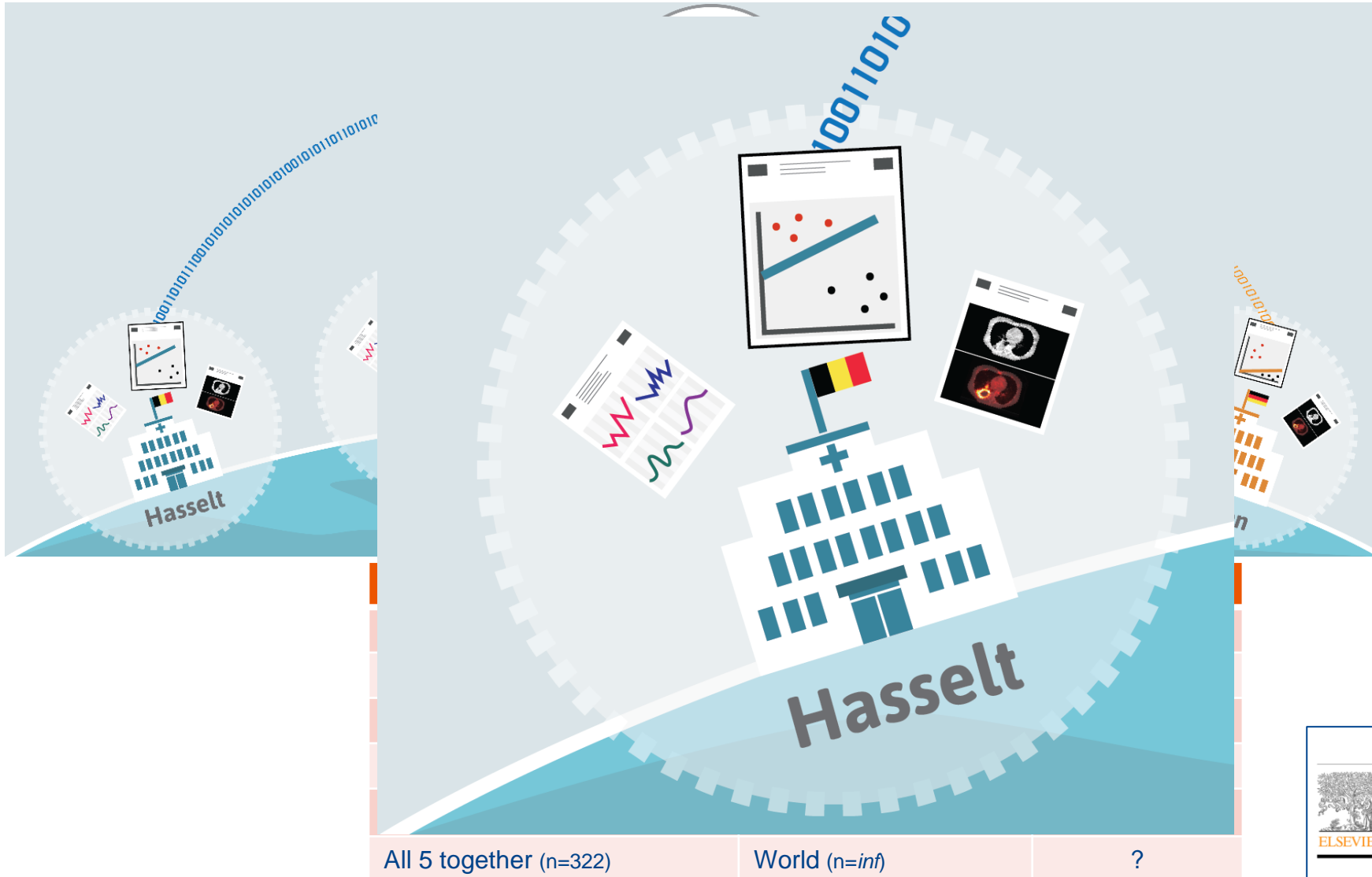
THURSDAY, APRIL 06, 2017

Posted by Brendan McMahan and Daniel Ramage, Research Scientists





# euroCAT example



# Federated Fair Data Infrastructures

- Toronto
- Boston
- Michigan
- Ottawa
- Philadelphia
- Tampa

- Poznan
- Oslo
- Bergen
- Odense
- Aachen
- Basel
- Zurich
- Nicosia
- Rome
- Aarhus

- Cardiff
- Manchester
- Leeds
- Sheffield
- Oxford
- Cambridge
- Hull

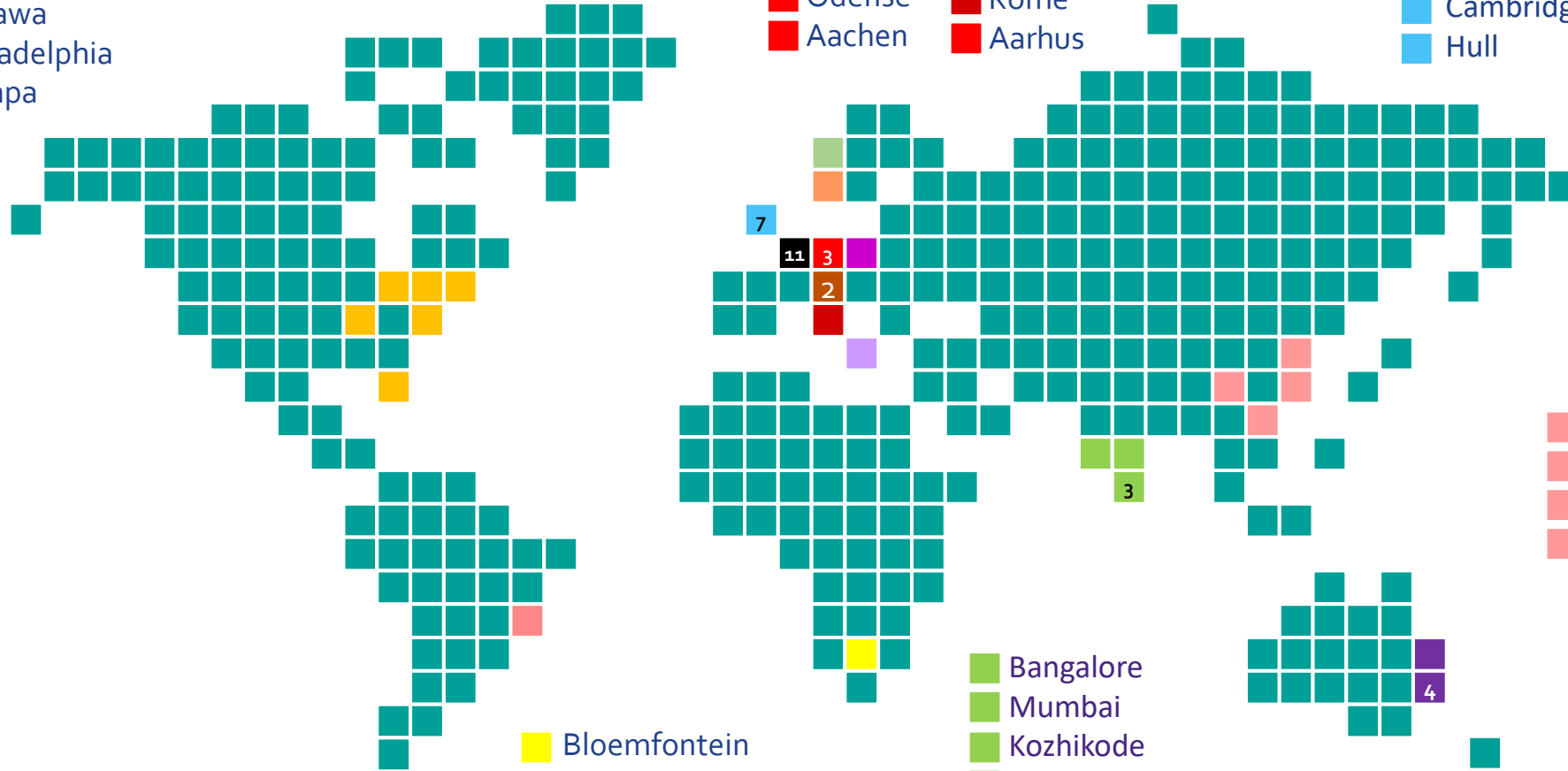
- Hasselt
- Eindhoven
- Maastricht
- Liege
- Nijmegen
- Amsterdam
- Rotterdam
- Zwolle
- Groningen
- Tilburg
- Leuven

- Shanghai
- Suining
- Tianjin
- Shantou

- Liverpool
- Wollongong
- Newcastle
- Westmead
- Coffs Harbour

- Bloemfontein
- Sao Paulo

- Bangalore
- Mumbai
- Kozhikode
- Hyderabad
- Vellore



# Applications

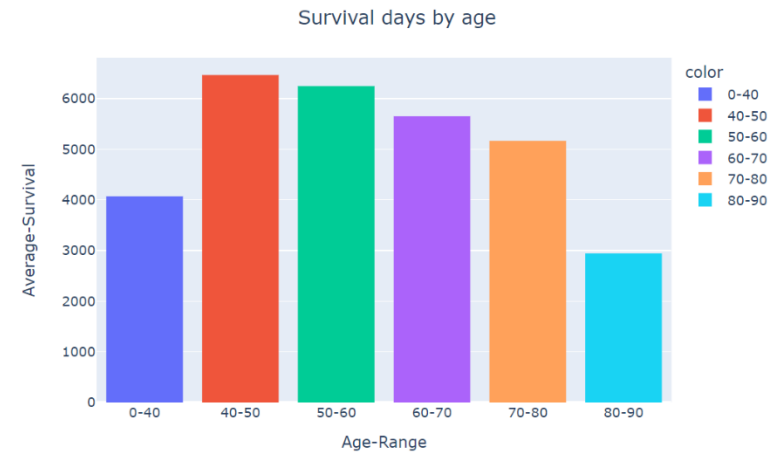
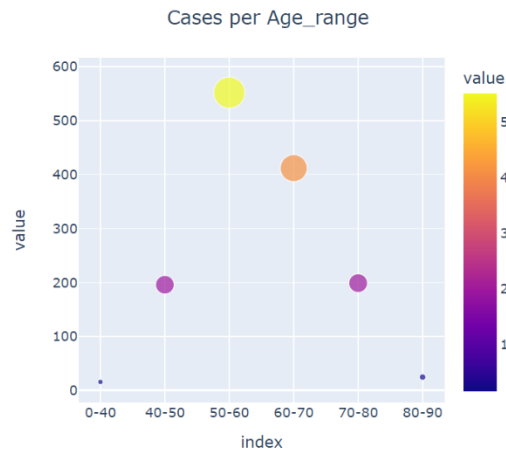
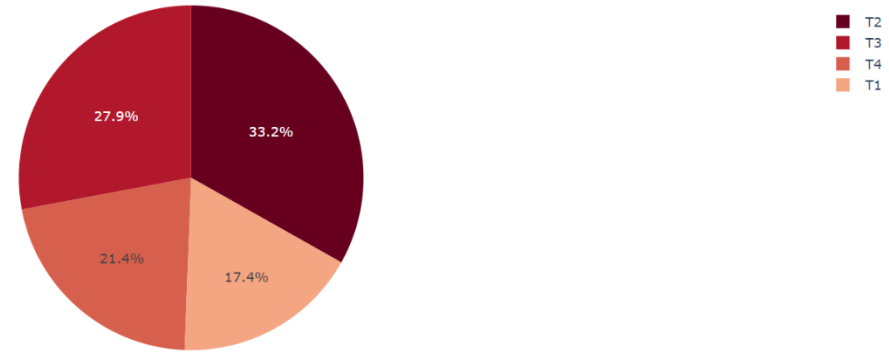


Nodes:

MAASTRO Private - Flyover Node  MAASTRO Public - Flyover Node  MD Anderson - Flyover Node  Montreal - Flyover Node  Toronto - Flyover Node

Choose an option:

T-stage



[CLICK TO DOWNLOAD DEMOGRAPHICS](#)



# Quality/Outcome Registries

Percentage radicaal bestraalde patiënten -met een WHO performance score van 0 of 1- met stadium III primair niet-kleincellig longcarcinoom\*, dat concurrent\*\* wordt behandeld.

42      54      77.8%      63.2%      Intern

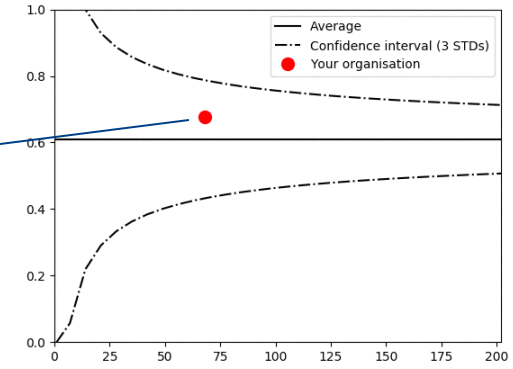
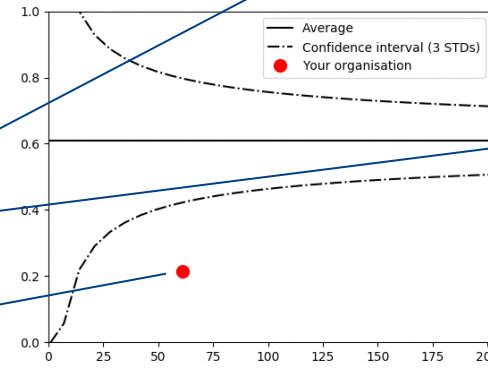
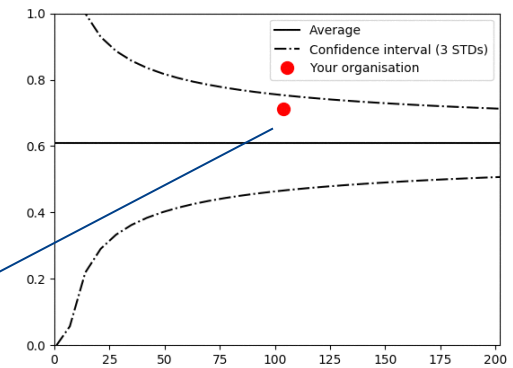
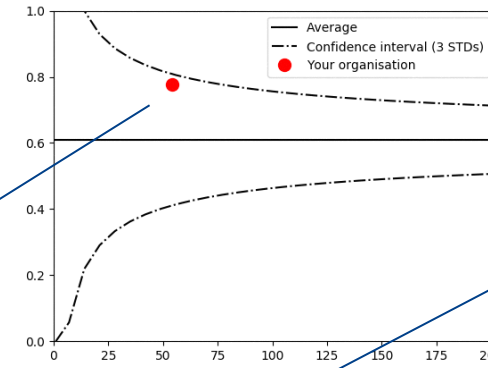
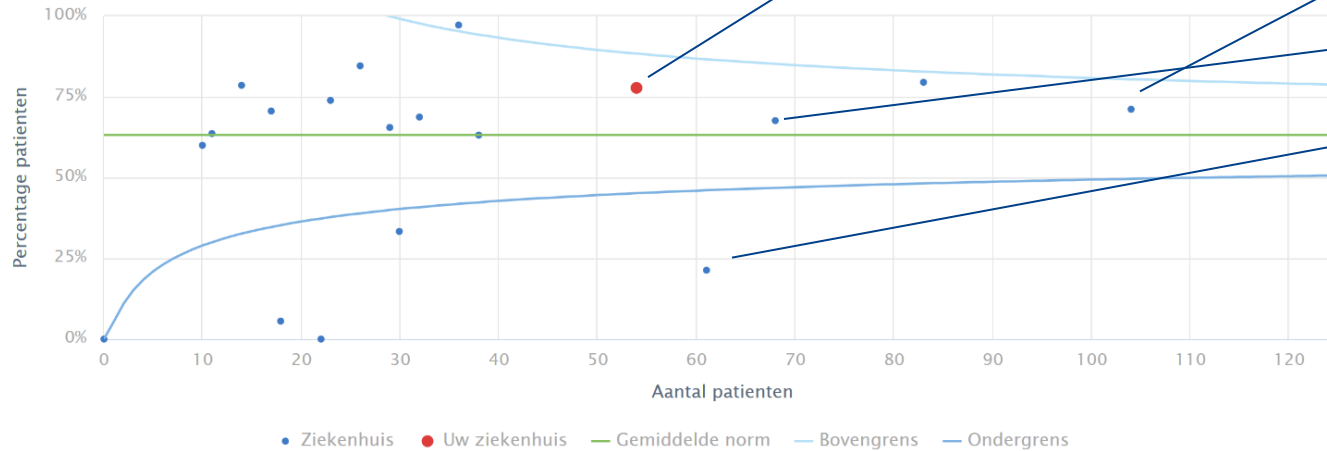
Teller:

Aantal patiënten dat een concurrent chemoradiotherapie behandelingsschema ondergaat

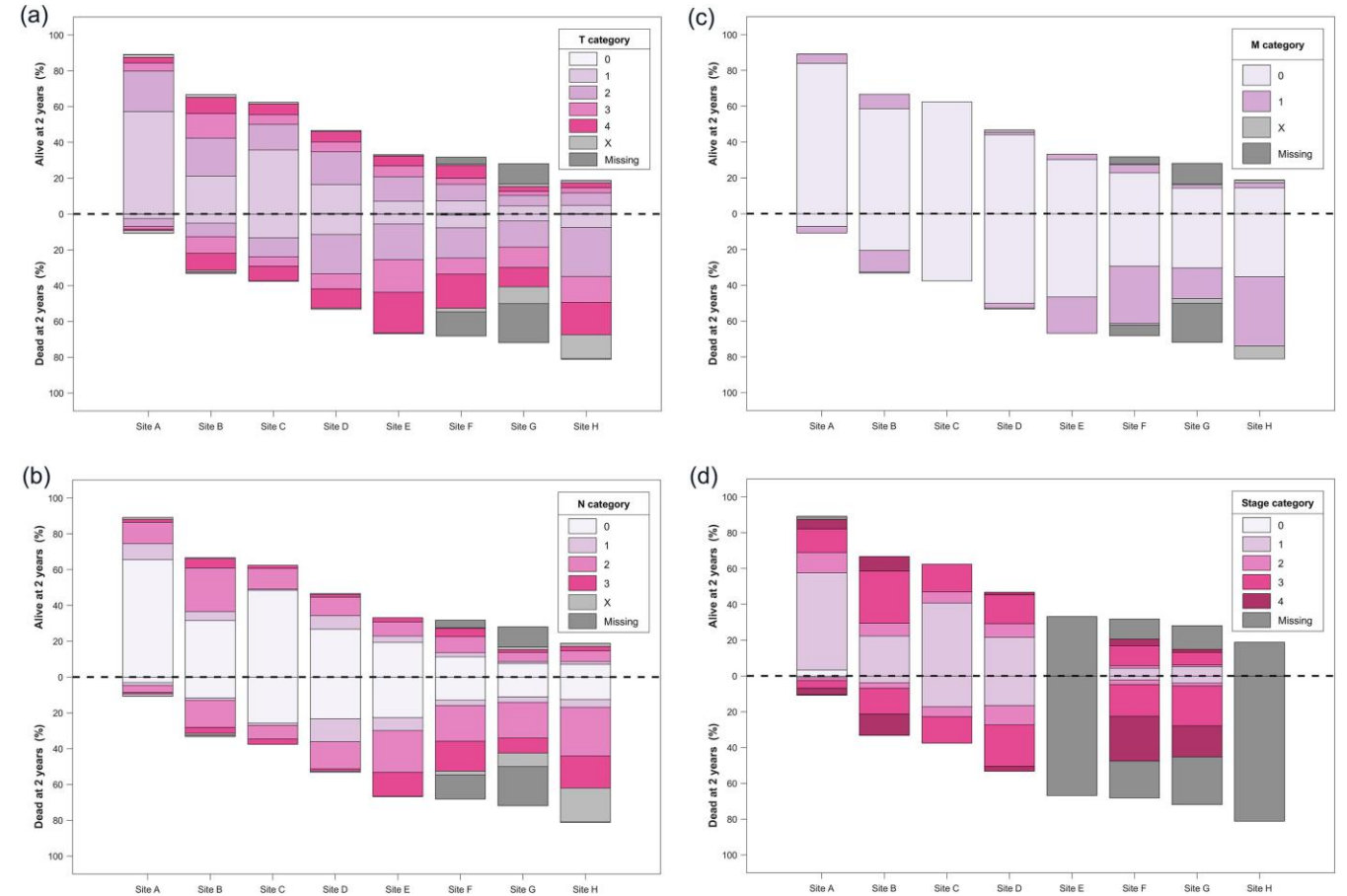
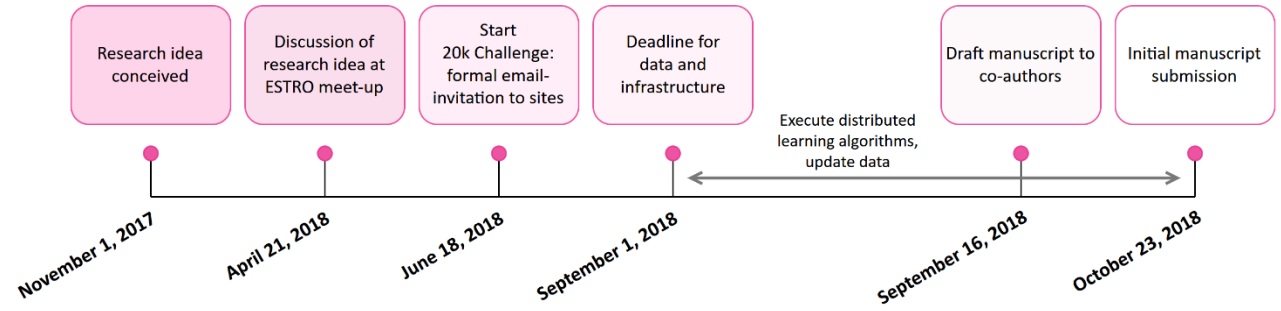
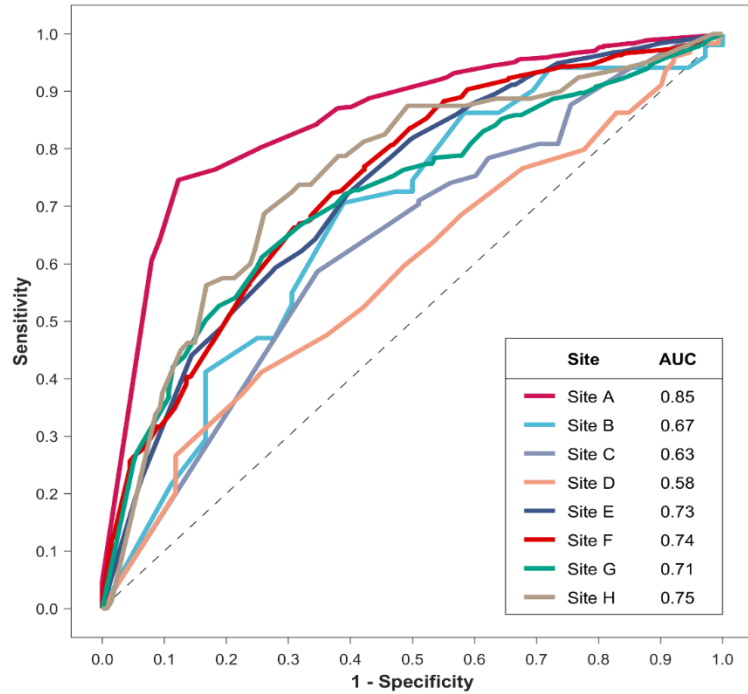
Noemer:

Aantal radicaal^ bestraalde patiënten met een WHO van 0-1, met een stadium III primair niet-kleincellig longcarcinoom jonger of gelijk aan 75 jaar

\*Niet-kleincellig longcarcinoom: Klinisch en/of pathologisch bewezen  
 \*\*Concurrent chemoradiotherapie: start radiotherapie ≤ 30 dagen na start chemotherapie  
 ^Radicaal bestraling = cumulatieve dosis van 40 tot 100 Gy



# 20k example



ORIGINAL ARTICLE | VOLUME 144, P189-200, MARCH 01, 2020

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

Timo M. Deist<sup>1</sup> • Frank J.W.M. Dankers<sup>1</sup> • Priyanka Ojha • ... Gareth Price<sup>2</sup> • Philippe Lambin<sup>2</sup> • Andre Dekker<sup>2</sup>

Open Access • Published: January 05, 2020 • DOI: <https://doi.org/10.1016/j.radonc.2019.11.019>

• Fully federated – no data sharing

## Example – 20k challenge

- Amsterdam, Cardiff, Maastricht , Manchester, Nijmegen, Rome, Rotterdam, Shanghai

Site	Available patients	Modelling cohort patient counts (complete cases, 1978–2015)						Model performance					
		Before imputation			After imputation			Training		Validation		Calibration-in-the-large	Calibration slope
		Training	Validation	Total	Training	Validation	Total	AUC	95%-CI	AUC	95%-CI		
Site A	5214	1050	3024	4074	1084	3058	4142	0.79	[0.75, 0.82]	0.85	[0.83, 0.87]	2.39	1.09
Site B	706	203	87	290	204	87	291	0.71	[0.62, 0.77]	0.67	[0.54, 0.78]	1.04	0.62
Site C	829	390	260	650	390	260	650	0.62	[0.57, 0.67]	0.63	[0.57, 0.69]	0.36	0.59
Site D	785	398	276	674	398	276	674	0.61	[0.55, 0.66]	0.58	[0.51, 0.64]	0.07	0.40
Site E	6211	0	0	0	2265	2458	4723	0.70	[0.68, 0.72]	0.73	[0.70, 0.75]	−0.09	0.85
Site F	4110	1165	520	1685	1906	1017	2923	0.73	[0.71, 0.76]	0.74	[0.71, 0.77]	0.20	0.96
Site G	16,260	6414	873	7287	6803	889	7692	0.74	[0.73, 0.75]	0.71	[0.68, 0.75]	0.02	0.75
Site H	2975	0	0	0	1760	348	2108	0.74	[0.71, 0.77]	0.75	[0.68, 0.80]	−0.43	0.76
<b>Total</b>	<b>37,090</b>	<b>9620</b>	<b>5040</b>	<b>14,660</b>	<b>14,810</b>	<b>8393</b>	<b>23,203</b>						

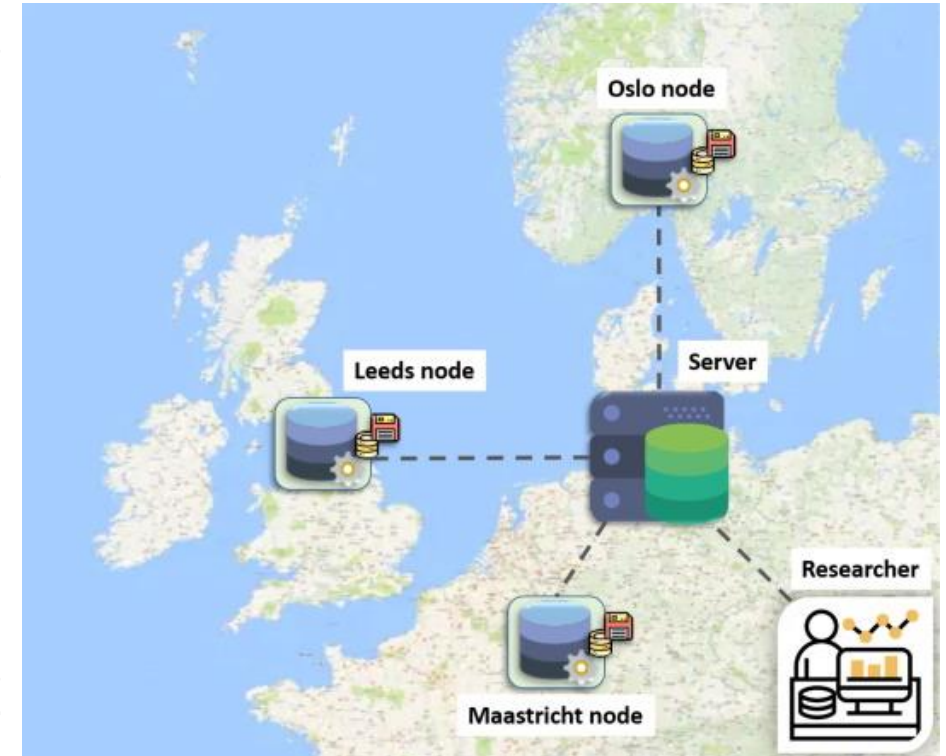
Deist, T. M. *et al.* Distributed learning on 20 000+ lung cancer patients – The Personal Health Train. *Radiotherapy and Oncology* **144**, 189–200 (2020).

# Rare cancers

atomCAT (n=281 from 3 sites) -> atomCAT2 (n=??? from 18 sites)

## \* Predicting outcomes in anal cancer patients using multi-center data and federated learning

Training nodes	MAASTRO Oslo	Leeds Oslo	Leeds MAASTRO
Validation node	Leeds	MAASTRO	Oslo
High risk disease (compared to low risk disease)	2.52 (0.93–6.78)	1.96 (0.68–5.67)	1.85 (0.71–4.86)
Male sex (compared to female sex)	3.59 (1.55–8.33)	3.83 (1.57–9.37)	2.12 (0.92–4.90)
Age at the start of RT	1.10 (0.74–1.64)	1.47 (0.99–2.17)	1.48 (1.05–2.10)
Primary tumour GTV	1.04 (1.00–1.08)	1.08 (1.03–1.13)	1.07 (1.03–1.11)
Primary tumour dose (EQD2)	0.97 (0.46–2.04)	0.35 (0.14–0.87)	0.97 (0.59–1.59)
Validation c-index	0.70	0.73	0.68
<b>Cox Model trained on all data :</b>	<b>0.72</b>	<b>0.74</b>	<b>0.70</b>



\* Radiotherapy and Oncology (2021) v159 p183-189,  
<https://doi.org/10.1016/j.radonc.2021.03.013>





# Distributed learning - making data research easier or just distributing the trouble?

*Ane Appelt*

*Associate Professor & Medical Physicist  
University of Leeds, UK*

*PBDW 2022*

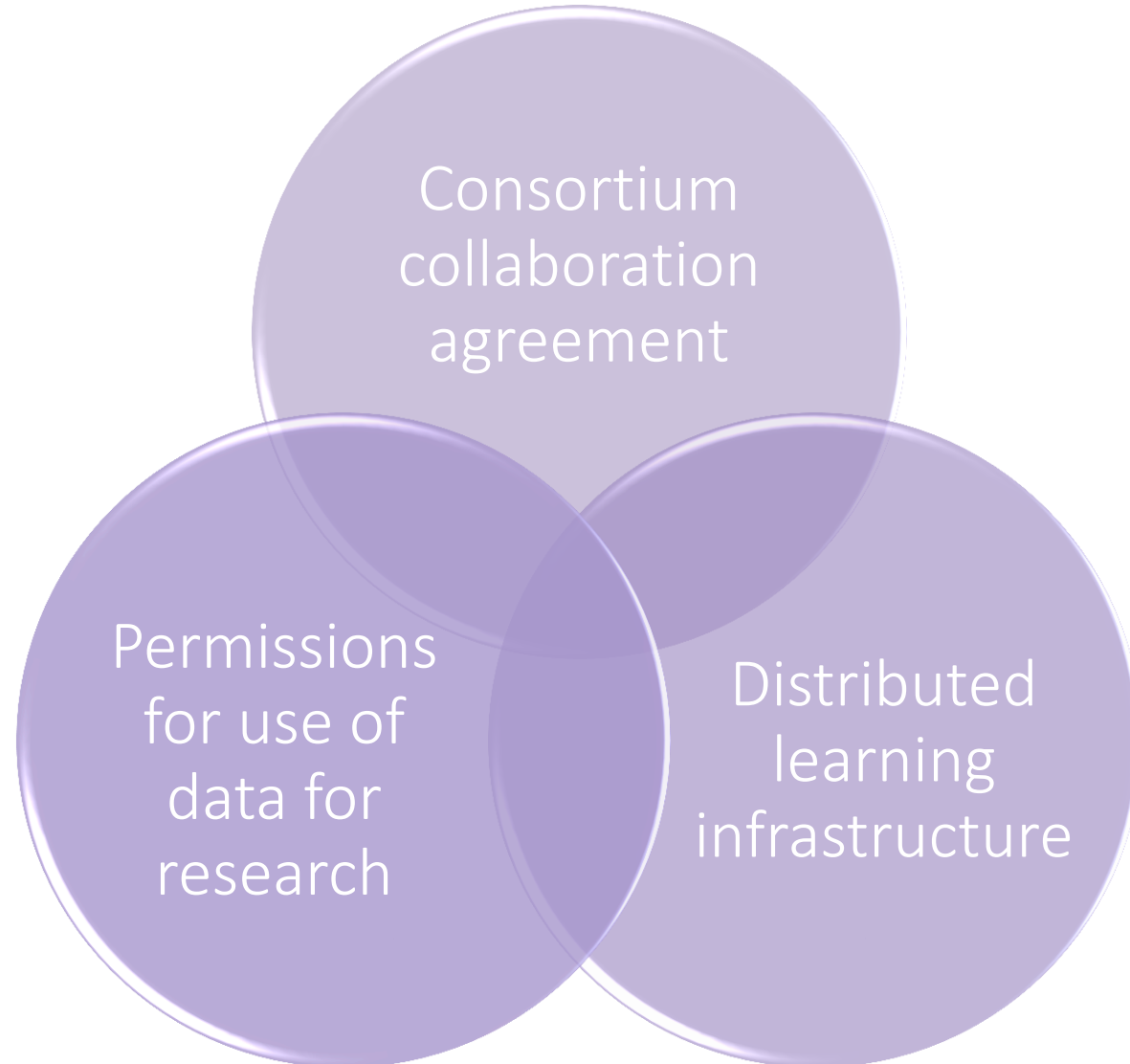


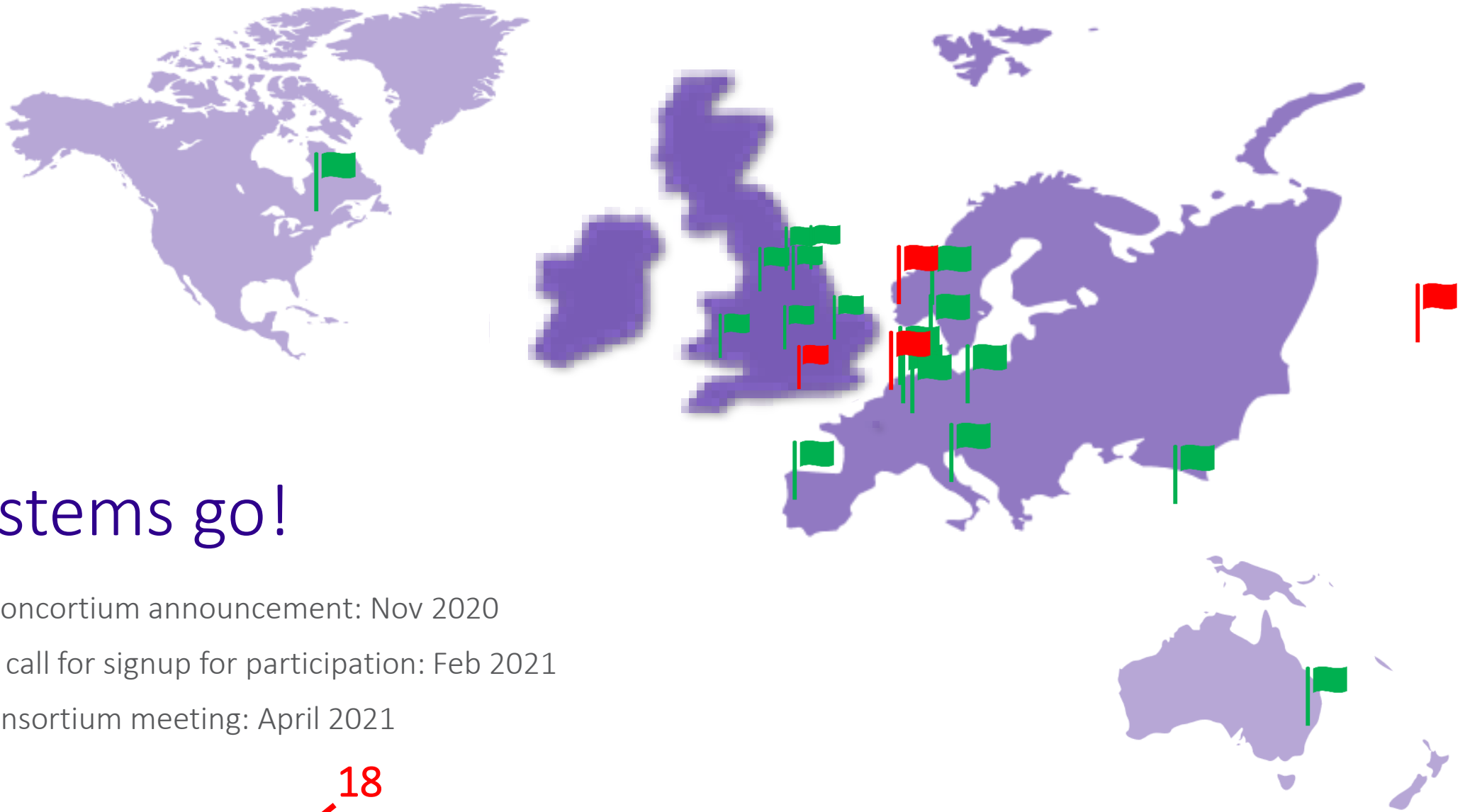
CANCER  
RESEARCH  
UK

RADNET  
LEEDS



# EACH CENTRE HAVE TO ENGAGE IN





## All systems go!

- Initial consortium announcement: Nov 2020
- Formal call for signup for participation: Feb 2021
- First consortium meeting: April 2021

Expression of interest: ~~24~~<sup>18</sup> centres



# CONSORTIUM COLLABORATION AGREEMENT

- First draft for centres – April 2021  
(additional reminders in Sept and Dec)
  - “Please engage your local legal teams”
- Full version for Leeds legal review – Dec 2021 – no substantial changes
- “Final” version circulated for consortium review: 24th Feb
- Deadline for “final” comments: 1st April 2022





“You want to install unknown software and open a network port on our hospital IT system? You got to be kidding me!”


# LEARNING POINTS

- 
- Local collection and curation of data still one of the largest barriers
  - Ethics of using data for research also an issue
  - And getting busy clinicians to work with their legal teams is almost impossible

# atomCAT2

1. Leeds Teaching Hospitals NHS Trust, Leeds, UK
2. Oslo University Hospital, Oslo, Norway
3. Maastricht University Medical Centre+, Maastricht, Netherlands
4. Ottawa Hospital and the University of Ottawa, Ottawa, Canada
5. Weston Park Hospital, Sheffield, UK
6. Policlinico Universitario A. Gemelli IRCCS, Università Cattolica S. Cuore, Rome, Italy
7. Ingham Research Institute and Liverpool Hospital, Liverpool, New South Wales, Australia
8. Addenbrooke's Hospital, Cambridge, UK
9. The Christie NHS Foundation Trust, Manchester, UK
10. RWTH Aachen University Medical Centre, Aachen, Germany
11. Oxford University Hospitals NHS Foundation Trust, Oxford, UK
12. Champalimaud Foundation, Lisbon, Portugal.
13. Greater Poland Cancer Center, Poznan, Poland
14. Hull University Teaching Hospitals NHS Trust, Hull, UK
15. Velindre University NHS Trust, Cardiff, UK.
16. Aarhus University Hospital, Aarhus, Denmark
17. The Netherlands Cancer Institute-Antoni van Leeuwenhoek (NKI-AVL), Amsterdam, The Netherlands
18. Bank of Cyprus Oncology Centre, Nicosia, Cyprus

Re: atomCAT2 analysis: Wednesday 2nd of November

 Stelios Theophanous [RPG] <umsth@leeds.ac.uk>

To: Ane Appelt; LILLEY, John (LEEDS TEACHING HOSPITALS NHS TRUST); eirik.malinen; perivarlonne; Leonard Wee; Ananya Choudhury; **+53 others**

Wed 02-Nov-22 09:34

If there are problems with how this message is displayed, click here to view it in a web browser.

Start your reply all with:    Feedback

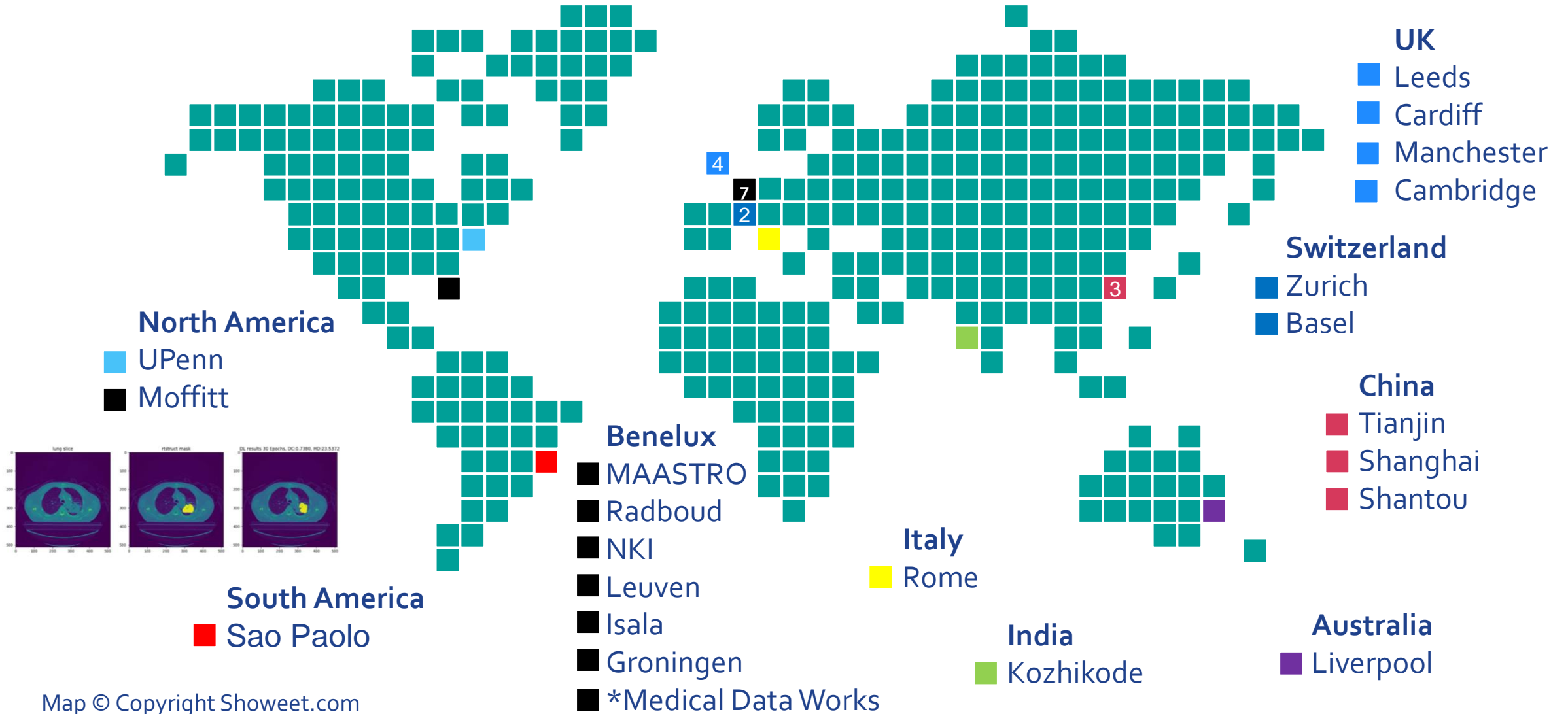
Good morning all,

We will be starting the atomCAT2 analysis in about 30 minutes. If possible, could you please start your nodes and let me know once you have done so?

During the analysis, I will be in the Teams meeting:

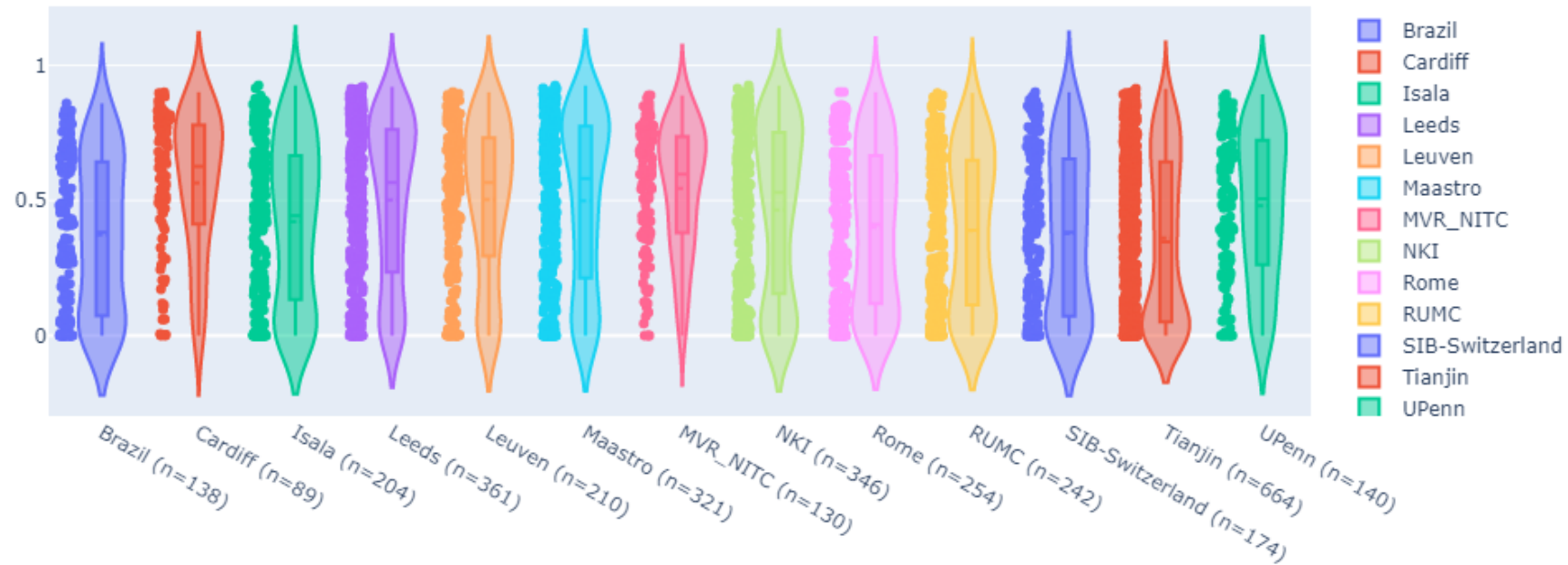
[https://teams.microsoft.com/l/meetup-join/19%3ameeting\\_MGI4ZDIkN2ItNzdhNS00Njc5LWJkMGtZDkxO](https://teams.microsoft.com/l/meetup-join/19%3ameeting_MGI4ZDIkN2ItNzdhNS00Njc5LWJkMGtZDkxO)

# Federated Deep Learning– ARGOS



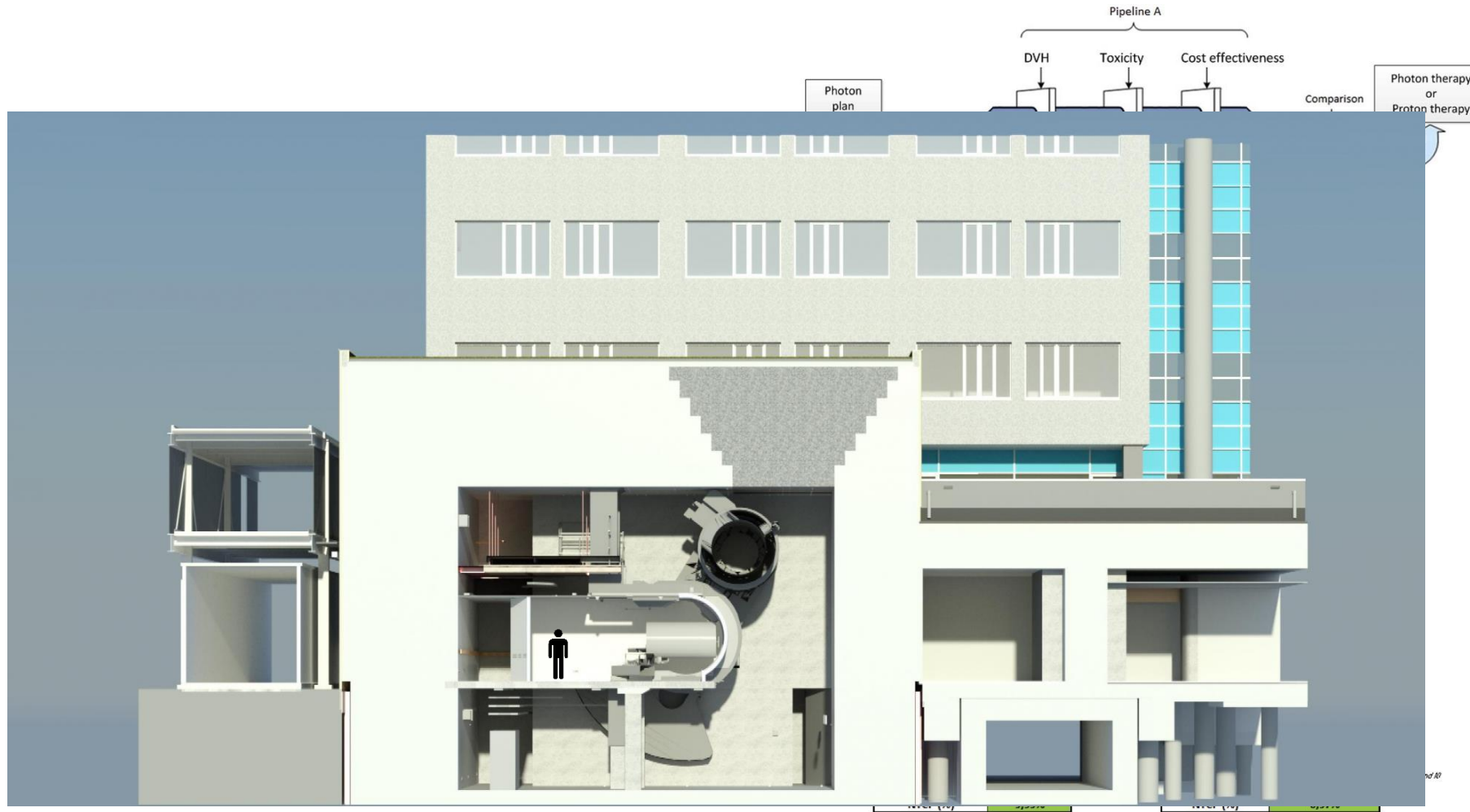
# AI for Lung GTV delineation – Oct 11, 2022

Mean Dice





# RWE/Model Based Indications - PROTRAIT



DELTA NTCP (%)	3,02%
RESULT	PROTON

# The value of federated data networks

- Only RCT based evidence is not feasible, we need to have complementary evidence.
- Real world data can generate evidence, but centralizing data is not doable
- FAIR federated data infrastructures offer a solution to collaborate
- Cautions
  - Causality | Culture | Legal | Quality | Ethics



# Acknowledgements

## Netherlands

MAASTRO, Maastricht, Netherlands  
 Radboudumc, Nijmegen, Netherlands  
 Erasmus MC, Rotterdam, Netherlands  
 Leiden UMC, Leiden, Netherlands  
 Elizabeth Twee Steden Ziekenhuis, Tilburg, 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

Aarhus Universitetshospital, Aarhus, Denmark

Bank of Cyprus Oncology Center, Nicosia, Cyprus  
 Weston Park Hospital, Sheffield, UK  
 Hull University Teaching Hospitals NHS Trust, Hull, UK  
 Addenbrookes' Hospital, Cambridge, UK

Oxford University Hospitals NHS Foundation Trust, Oxford, UK  
 Haukeland University Hospital, Bergen, Norway

## Africa

University of the Free State, Bloemfontein, South Africa

## Asia

Fudan Cancer Center, Shanghai, China  
 CDAC, Pune, India  
 Tata Memorial, Mumbai, India  
 Suining Central Hospital, Suining, China  
 HGC Oncology, Bangalore, India  
 MVRCC&NITC, Calicut, Kerala, India  
 Apollo Hospitals, Hyderabad, India  
 CMC Vellore, Vellore, India  
 Tianjin Medical University, Tianjin, China  
 Cancer Hospital of Shantou University, Shantou, China

## North America

RTOG, Philadelphia, PA, USA  
 MGH, BWH, Harvard, Boston, MA, USA  
 University of Michigan, Ann Arbor, USA  
 Princess Margaret CC, Canada  
 Ottawa Hospital Research Institute, Ottawa, 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



## Clinical Data Science research aims

1. Get access to all data of all people in the world
2. Learn personalized health prediction models from data
3. Apply prediction models to improve health

Cancer, Alzheimer's, Cardiovascular disease, Diabetes, Heart Failure, Parkinson's, Irritable Bowel Disease, Orthopedic Surgery, Rheumatoid Arthritis, Pediatric Surgery, Balance disorders, Hip dysplasia

Thank you for your attention