PHILIPS

Philips Guest Lecture

Data Science: data acquisition and analysis

Ymke de Jong, Tim Hulsen

24 OCT 2023

innovation you

Agenda

1. Intro Philips & data&AI
2. Data Management at Philips
3. Data collection example e/MTIC
4. Data collection example ReIMAGINE
5. Close-up remarks
6. Q&A

© Koninklijke Philips N.V.

Who are we

Ymke de Jong

Study:
- Biomedical Engineering TU/e

Work:
- Data Scientist @ BPSolutions
- Data Scientist @ Achmea
- Clinical Data Scientist @ Philips
- **Technical Program manager Data&AI @ Philips**

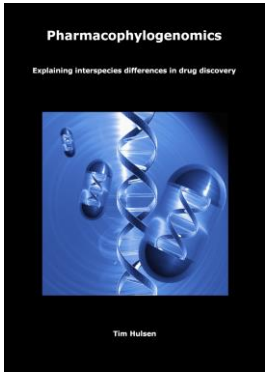
/ymke-de-jong

Tim Hulsen

Study:
- MSc in Biology @ Radboud

Work:
- Bioinformatician @ Organon
- PhD in Bioinformatics @ Radboud / Organon
- PostDoc in Bioinformatics @ Radboud / Schering-Plough
- **Senior Scientist @ Philips**
  (project manager, data scientist, data custodian)

/thulsen



© Koninklijke Philips N.V.

Royal Philips

© Koninklijke Philips N.V.

Philips, a born innovator

For over 130 years, we have been improving people's lives with a steady flow of ground-breaking innovations

Philips has
reinvented itself many times

01
Founded on innovation and entrepreneurship

02
Expanding beyond lighting

03
Global expansion post World War II

04
Diversified industrial conglomerate

05
Strategic portfolio choices sharpening focus

Philips 6.0
A focused leader in **health technology**
With enviable, established positions in the hospital and the home, Philips is entering this next phase in its history determined to extend its leadership in health technology.

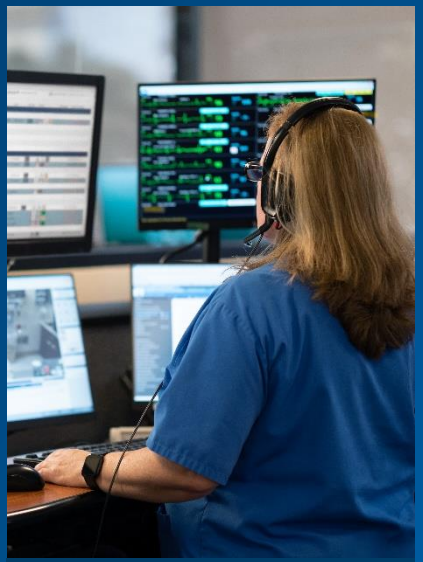
Healthcare

© Koninklijke Philips N.V.

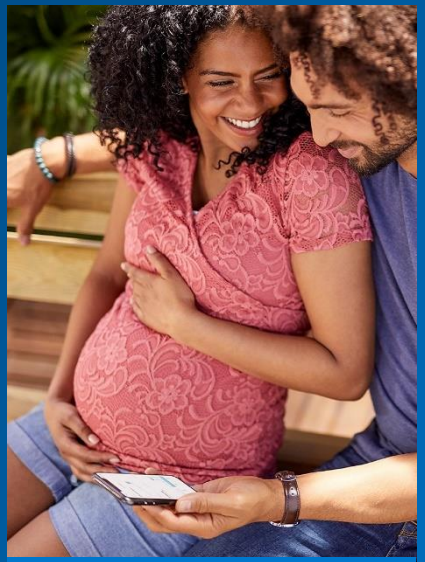
Healthcare providers face an unsustainable burden, while patient expectations of healthcare are changing

Populations are ageing and chronic diseases are on the rise

Growing staff shortages are causing burnout



Healthcare providers are seeking efficiencies to reduce cost



People are increasingly engaged in their own health

© Koninklijke Philips N.V.

Health is a fundamental human right

Yet half the world's population lacks access to the health services they need

Source: World Health Organization

Digital transformation is driving exponential growth of health data

- ✓ Medical imaging
- ✓ Patient monitoring
- ✓ Pathology
- ✓ Genomics
- ✓ Home monitoring
- ✓ Personal health tracking

By 2025, the amount of **healthcare data** is expected to **grow** by **36% per year**

Source: https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion

© Koninklijke Philips N.V.

“

WE ARE DROWNING IN INFORMATION WHILE STARVING FOR WISDOM

E.O Wilson

© Koninklijke Philips N.V.

Examples across the health continuum

© Koninklijke Philips N.V.

Identifying at-risk patients in critical care units

Using predictive algorithms for clinical decision support

Keeping a caring eye on patients during MR exams

Using touchless patient sensing and advanced algorithms to detect breathing

Increasing operational efficiency in MR

Minimizing unexpected downtime in case of an operational issue

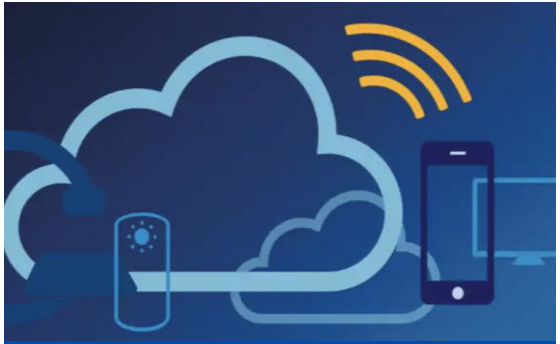
Empowering people to take care of their oral health

With personalized recommendations for better brushing habits

Data @ Philips

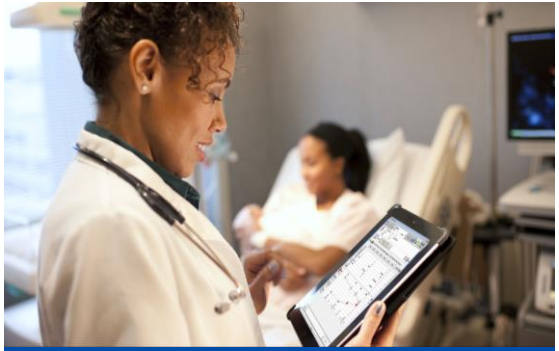
© Koninklijke Philips N.V.

Addressing data-related challenges by promoting data sharing, interoperability and next-generation AI methods



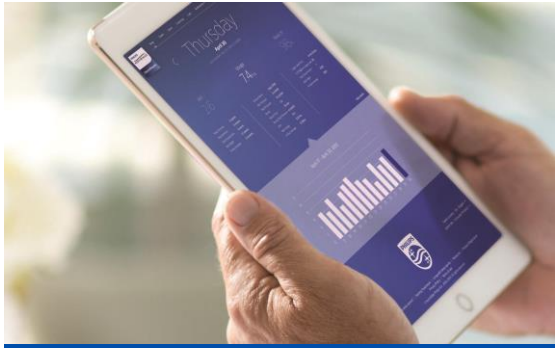
Philips HealthSuite cloud-based digital platform

We help unlock data from different systems and devices in and outside the hospital



Open data standards and interoperability

We enable healthcare providers to connect and integrate data in a meaningful way

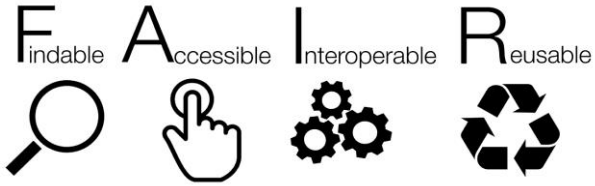


Pioneering next-generation AI methods

We are actively exploring the use of next-gen AI methods such as federated learning to address data-related challenges

© Koninklijke Philips N.V.

FAIR principles



The US National Institutes of Health is located in Bethesda, Maryland.

NIH ISSUES A SEISMIC MANDATE: SHARE DATA PUBLICLY

Policy could set a standard for research, scientists say, but they have questions about logistics and equity.

https://www.nature.com/articles/d41586-022-00402-1

Study finds the risks of sharing health care data are low

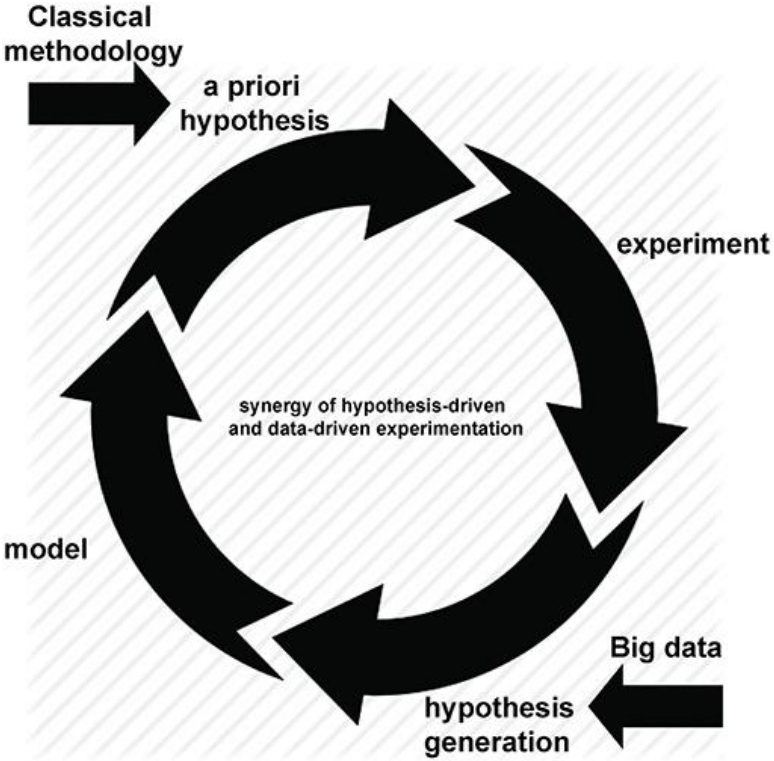
Greater availability of de-identified patient health data would enable better treatments and diagnostics, the researchers say.

Anne Trafton | MIT News Office
October 6, 2022

https://news.mit.edu/2022/patient-data-risks-low-1006

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/

Big data: hypothesis-driven & data-driven research



- "A priori hypotheses are no longer necessary" vs. "Data-driven approaches are just a distraction"
- No! Hypotheses generate data, data generate hypotheses.
- Hypothesis-generating approaches are not only synergistic with traditional methods; they are dependent upon them: after all, once generated, a hypothesis must be tested. In this way, Big data analyses can be used to ask novel questions, with conventional experimental techniques remaining just as relevant for testing them.

From: Hulsen T. et al. From Big Data to Precision Medicine.
Front. Med. (2019) 6: 34 (CC BY 4.0 – unmodified).
https://www.frontiersin.org/articles/10.3389/fmed.2019.00034/full

Data management

- There are many opportunities to exploit AI to improve healthcare, but these need proper data management.

- Data management = a fast-growing scientific field

- Artificial Intelligence (AI): garbage in, garbage out ('GIGO'). In other words: even the most fancy AI algorithm will produce bad results if the data are 'bad':
  - Many missing data ('N/As')
  - Badly curated data
  - Biased data
  - Missing / low quality metadata (see 'FAIR Principles')
  - etc.

- Data management is related to terms such as:
  - Data stewardship
  - Data custody
  - Data integration
  - Data curation

- Job of data management is to make sure that no garbage gets in, to go 'from data to intelligence'

- Enable data sharing, within context of privacy and intellectual property

From: Hulsen, T., Petkovic, M. et al. AI in Healthcare: From Data to Intelligence *Front. Artif. Intell.* (2022) 5: 909391. https://www.frontiersin.org/articles/10.3389/frai.2022.909391/full

From: Hulsen, T. Sharing Is Caring - Data Sharing Initiatives in Healthcare *Int. J. Environ. Res. Public Health* (2020) 17 (9): 3046. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246891/

Ten commandments of translational research informatics

Translational research informatics = data management supporting 'translational' (= from bench to bedside) informatics

These ten commandments help to make the translational research project a success:

1) Create a separate data management work package
2) Reserve time and money for data entry
3) Define all data fields up front with the help of data analysis experts
4) Make clear arrangements about data access
5) Agree about de-identification and anonymization
6) Reuse existing software where possible
7) Make newly created software reusable
8) Adhere to the FAIR Guiding Principles
9) Make sure that successors are being instructed correctly
10) Make it sustainable: what happens after the project?

From: Hulsen, T. The ten commandments of translational research informatics. *Data Science* (2019) 2: 341–352 (CC BY 4.0 – unmodified).
https://content.iospress.com/articles/data-science/ds190020

Exemplary initiatives Philips is involved

nature › scientific data › data descriptors › article

Open Access | Published: 24 May 2016

MIMIC-III, a freely accessible critical care database

Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi & Roger G. Mark

Scientific Data 3, Article number: 160035 (2016) | Cite this article

149k Accesses | 1775 Citations | 81 Altmetric | Metrics

Abstract

MIMIC-III ('Medical Information Mart for Intensive Care') is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital. Data includes vital signs, medications, laboratory measurements, observations

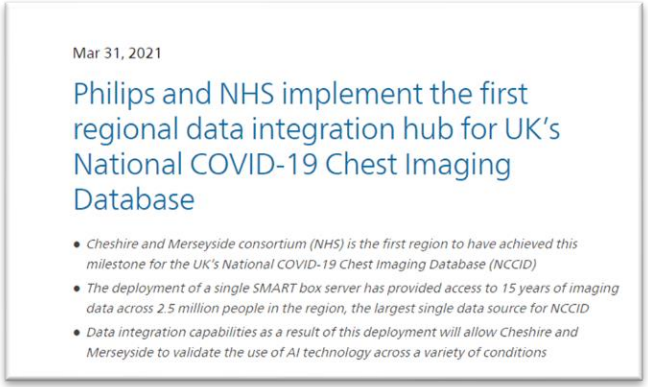
Article | Open access | Published: 24 June 2023

A guide to sharing open healthcare data under the General Data Protection Regulation

Jip W. T. M. de Kok, Miguel Á. Armengol de la Hoz, Ymke de Jong, Véronique Brokke, Paul W. G. Elbers, Patrick Thoral, Alejandro Castillejo, Tomás Trenor, Jose M. Castellano, Alberto E. Bronchalo, Tobias M. Merz, Martin Faltys, Collaborator group, Iwan C. C. van der Horst, Minnan Xu, Leo Anthony Celi, Bas C. T. van Bussel & Xavier Borrat

Scientific Data 10, Article number: 404 (2023) | Cite this article

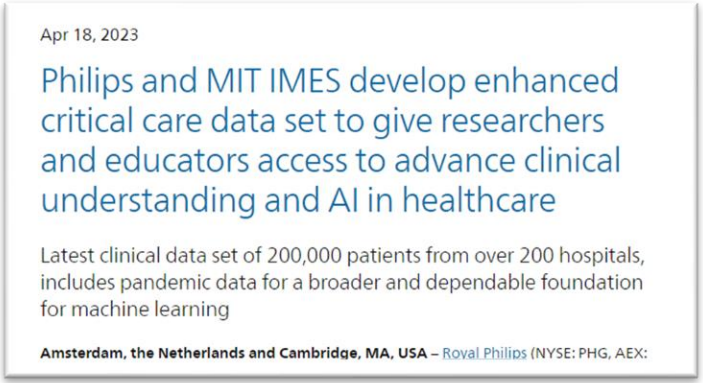
5887 Accesses | 31 Altmetric | Metrics



Mar 31, 2021

Philips and NHS implement the first regional data integration hub for UK's National COVID-19 Chest Imaging Database

- *Cheshire and Merseyside consortium (NHS) is the first region to have achieved this milestone for the UK's National COVID-19 Chest Imaging Database (NCCID)*
- *The deployment of a single SMART box server has provided access to 15 years of imaging data across 2.5 million people in the region, the largest single data source for NCCID*
- *Data integration capabilities as a result of this deployment will allow Cheshire and Merseyside to validate the use of AI technology across a variety of conditions*



Apr 18, 2023

Philips and MIT IMES develop enhanced critical care data set to give researchers and educators access to advance clinical understanding and AI in healthcare

Latest clinical data set of 200,000 patients from over 200 hospitals, includes pandemic data for a broader and dependable foundation for machine learning

Amsterdam, the Netherlands and Cambridge, MA, USA – Royal Philips (NYSE: PHG, AEX:

From:
https://pubmed.ncbi.nlm.nih.gov/27219127/
https://www.nature.com/articles/s41597-023-02256-2
https://www.philips.com/a-w/about/news/archive/standard/news/press/2021/20210331-philips-and-nhs-implement-the-first-regional-data-integration-hub-for-uk-s-national-covid-19-chest-imaging-database.html
https://www.philips.com/a-w/about/news/archive/standard/news/press/2023/20230418-philips-and-mit-imes-develop-enhanced-critical-care-data-set-to-give-researchers-and-educators-access-to-advance-clinical-understanding-and-ai-in-healthcare.html

ReIMAGINE

© Koninklijke Philips N.V.

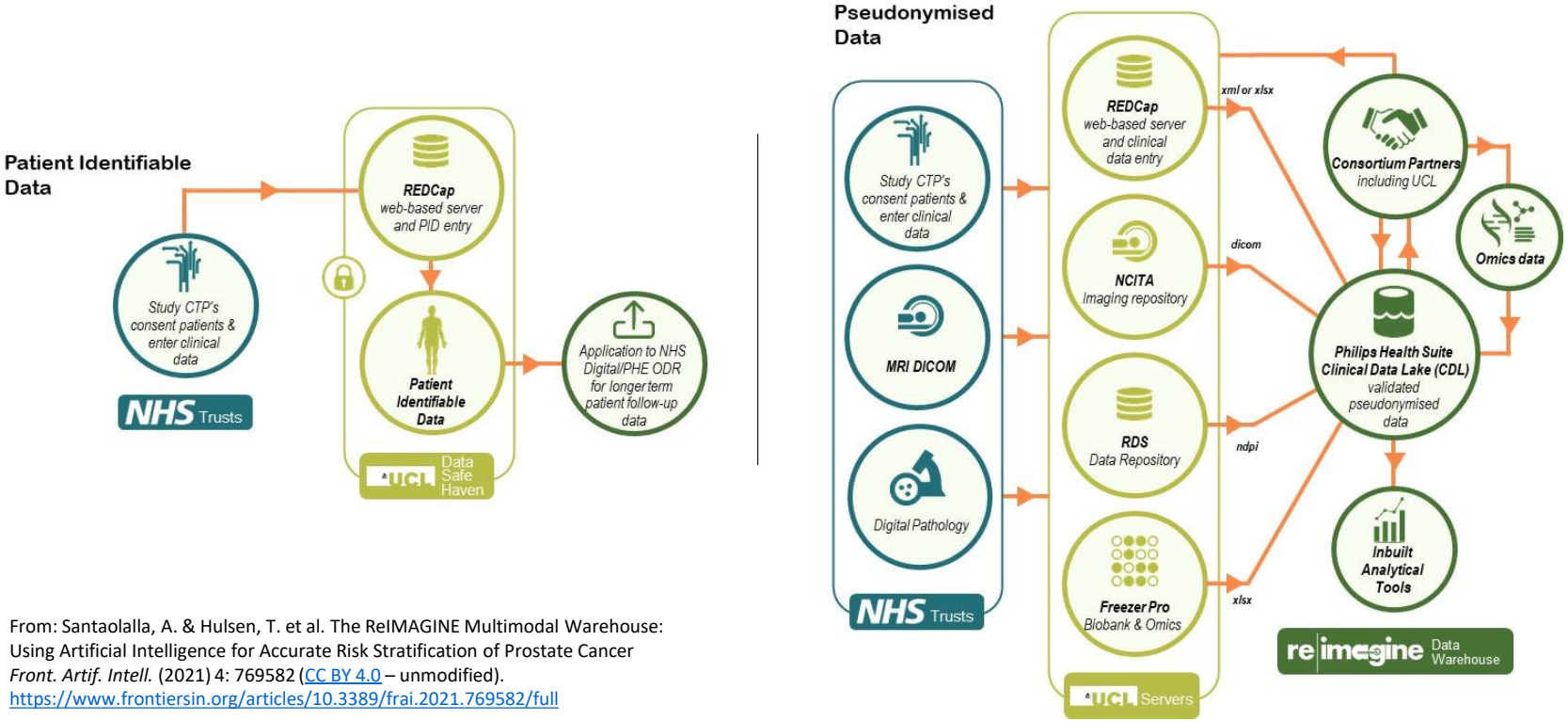
Case study: ReIMAGINE: introduction

- ReIMAGINE - correcting five decades of error through enabling image-based risk stratification of localized PCa
- The ReIMAGINE study will test for the first time if MRI scans can be used for population screening to detect prostate cancer more accurately. The current prostate-specific antigen (PSA) test is considered too unreliable for population screening, and the scientists will study if MRI could be used to screen men to pick up cancers earlier and more reliably, and help save lives.
- They will also study whether, combined with cutting-edge techniques such as genomics and machine learning, MRI scans can replace prostate biopsies. Prostate cancer patients advising the study say they are particularly excited by the prospect of large reduction in biopsies, as they have serious side effects in the majority of patients, which include pain, bleeding, infections leading to sepsis, and urine retention.
- PI: Mark Emberton (UCL). Timeline: August 2018 – June 2023 UCL KING'S College LONDON PHILIPS a.o.
- Two patient cohorts: WS1 'risk' (n=1000) and WS2 'screening' (n=300). Data generated: clinical, imaging, pathology and omics, incl. followup data. See https://www.reimagine-pca.org/ re|imagine

WS1: Marsden T. et al. The ReIMAGINE prostate cancer risk study protocol: A prospective cohort study in men with a suspicion of prostate cancer who are referred onto an MRI-based diagnostic pathway with donation of tissue, blood and urine for biomarker analyses. *PLoS One* (2022) 17 (2): e0259672. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0259672

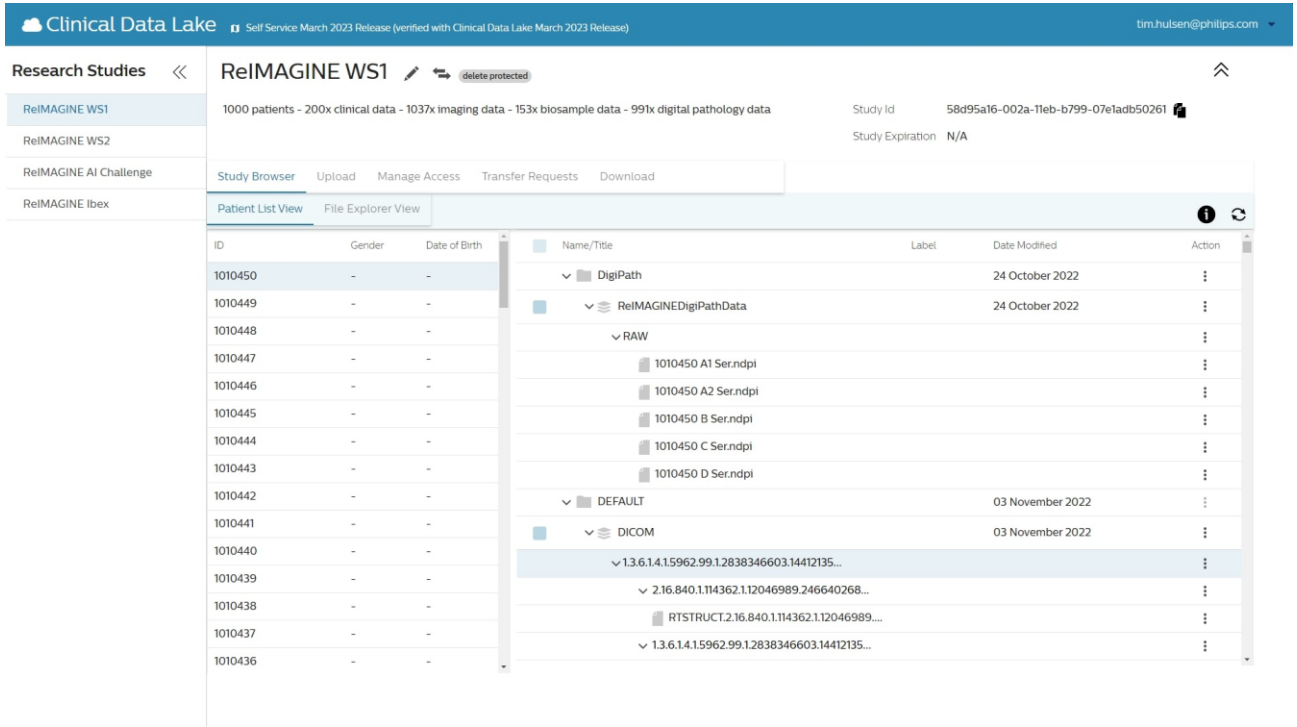
WS2: Marsden T. et al. ReIMAGINE Prostate Cancer Screening Study: protocol for a single-centre feasibility study inviting men for prostate cancer screening using MRI. *BMJ Open* (2021) 11 (9): e048144. https://bmjopen.bmj.com/content/11/9/e048144.long

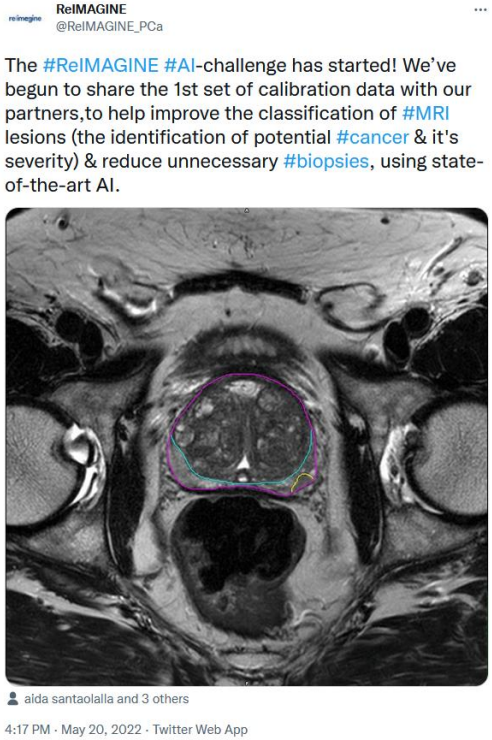
Case study: ReIMAGINE: data management



From: Santaolalla, A. & Hulsen, T. et al. The ReIMAGINE Multimodal Warehouse: Using Artificial Intelligence for Accurate Risk Stratification of Prostate Cancer *Front. Artif. Intell.* (2021) 4: 769582 (CC BY 4.0 – unmodified). https://www.frontiersin.org/articles/10.3389/frai.2021.769582/full

Case study: ReIMAGINE: Clinical Data Lake (CDL)





From: Santaolalla, A. & Hulsen, T. et al. The ReIMAGINE Multimodal Warehouse: Using Artificial Intelligence for Accurate Risk Stratification of Prostate Cancer *Front. Artif. Intell.* (2021) 4: 769582 (CC BY 4.0 – unmodified). https://www.frontiersin.org/articles/10.3389/frai.2021.769582/full

From: https://twitter.com/ReIMAGINE_PCa/status/1527654771280207874

Example data collection e/MTIC

If we have time!

© Koninklijke Philips N.V.

e/MTIC – Eindhoven MedTech Innovation Center

Key Take Aways

© Koninklijke Philips N.V.

Key Take Aways

1. **Data as foundation for AI -** Data management (as well as data curation, data stewardship, etc.) needs to be taken into account when trying to do data science. If not, problems such as low data quality or limited reproducibility might arise.

1. **Multidisciplinary –** Clinical expertise is utmost important for AI in healthcare and privacy is a cornerstone of the data science field. Make sure you understand and work with it!

2. **Teamwork** - The TRICA Healthdot (with e/MTIC) and the ReIMAGINE (with UCL) projects are examples of collaborations between hospitals, academia and industry in which big data are used to create applications that can be used in the clinic.

Data Science in Healthcare is not about the perfect algorithm.

It is about designing and developing data driven solutions that can be adopted, embraced and adopted by the healthcare providers and patients towards including full solution lifecycle.

© Koninklijke Philips N.V.

PHILIPS