Project title: Accelerating the lab to market transition of AI tools for cancer management
Grant Agreement: 952172
Call identifier: H2020-SC1-FA-DTS-2019-1
Topic: DT-TDS-05-2020 AI for Health Imaging

D3.1 ACCESSIBLE IMAGING HEALTH DATA MAP

Leader partner: Maastricht University, UM
Author(s): UM: Anke Wind, Manon Beuque, Henry Woodruff
UNIPI: Emanuele Neri, Lorenzo Tumminello
HULAFE: Luis Martí-Bonmatí, Ana Miguel Blanco

Work Package: WP3
Due date: Month 9
Actual delivery date: 31/05/2021
Type: Deliverable type R = Report
Dissemination level: PU = public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

Table of contents

Abbreviations

ADNI	Alzheimer's Disease Neuroimaging Initiative
Csv	Comma-separated values
DICOM	Digital Imaging and Communications in Medicine
DoD	Department of Defense
DOIs	Digital Object Identifier System
DTI	Diffusion tensor imaging
DXA	Dual-energy X-ray absorptiometry
EUSOMA	European Society of Breast Cancer Specialists
FAIR	Findable; Accessible; Interoperable; Reusable
FITBIR	Federal Interagency Traumatic Brain Injury Research
GDPR	General Data Protection Regulation
MRI	Magnetic resonance imaging
Nifti	Neuroimaging Informatics Technology Initiative
NIH	National Institutes of Health
NLM	National Library of Medicine
OAI	Osteoarthritis Initiative
PET	Positron emission tomography
SICAS	Study In China Admission System
SIREN	Stroke Investigative Research & Educational Network
SNP	single nucleotide polymorphisms
TBI	Traumatic Brain Injury
UMC	University Medical Centre
UMCG	University Medical Centre Groningen

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

Disclaimer

The opinions stated in this report reflect the opinions of the authors and not the opinion of the European Commission.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

1. Introduction

CHAIMELEON aims to develop a Distributed Data Repository interoperable with existing repositories, enabling secure sharing and reuse of anonymized imaging and clinical data, for AI developers to develop cancer management tools and solutions.

This deliverable gives an overview of existing initiatives in the form of an Imaging Health Data Map that gives the project the opportunity to learn from best practices in order to reach the following key objectives of the project:

O2: To design an EU-wide interoperable repository of health images, related clinical data and molecular data from pathologic and liquid biopsy samples, as a useful resource for the AI community to develop and test effective tools for improved cancer management applications.

O3: To set up the repository as a distributed infrastructure, building on existing initiatives at European, national, regional and individual centres levels.

Partners of the CHAIMELEON Consortium are composed of medical and technical experts, from both academia and industry. Different partners already work with or know of imaging repositories. This deliverable synthesizes information on these already existing imaging repositories.

In the first part of the document (section 2 Methods), we describe how we compiled this overview. We then move to presenting our findings in the Results section in two structured tables showing various aspects of the listed repositories.

In the Discussion, we describe which lessons can be learned that can be taken into account in the design of the CHAIMELEON repository and what conclusions can be drawn from this accessible Imaging Health Data Map.

2. Methods

2.1 Questionnaire

To generate the map we created a questionnaire using Google forms. In this questionnaire we asked members of the consortium to anonymously list any health data repositories that they knew. For each repository we asked them to list the following information each repository/silo:

- Name
- Link or other source of information

If known:

- Type of data included
- Is the metadata available
- Is the data available for download as is
- Access, who and how (for example by use of form)
- Data inclusion control, who decides which data is included?

A link to this questionnaire was sent to all consortium partners and shared again during the consortium meeting. Besides this questionnaire, we searched for additional repositories in google and through an overview by aylword.org[1] to come up with a comprehensive list.

2.2 Creating the map

To get a complete overview each listed database was checked to assess the factors needed for creating the map by two researchers. After internal revision an additional three datasets

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

were added to the list. The map consists of two parts, the first (3.1) being an overview consisting of a general description of the repository; type of data included; and the data inclusion control (if clearly stated). The second part of the map (3.2) consists of more specific information about each repository structured based on the FAIR (**F**indable; **A**ccessible; **I**nteroperable; **R**eusable) elements. The third map (3.3) includes repositories/biobanks that work send after an internal review. We had a brief review of these, but did not have time to completely review them and add those that were applicable to the first two maps. However, after our short review we did not find information that will change our conclusions and lessons learnt.
At all three maps for all repositories, parts that were unknown are left blank or, if clear a reason for not knowing the information is given.

2.3 Creation and maintenance of the Radiomics Imaging Archive (RIA)

In order to solve the unmet need for a disease agnostic (every type of disease) GDPR compliant platform to share imaging and associated data which is "AI ready", i.e., is fully curated to the needs of machine learning, we have designed and implemented the Radiomics Imaging Archive (RIA) https://www.radiomicsimagingarchive.eu/welcome. This serves as both a central repository for open source data as well as the storage for local projects which cannot be shared without a prior legal agreement such as a Data Transfer Agreement. The lessons learned from this project are also incorporated in the Discussion.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

3. Results

3.1 Overview repositories

Name (link)	Description	Data	Data inclusion control
PRIMAGE Repository[2] https://www.primageproject.eu/	PRIMAGE platform implements the latest advancement of in-silico imaging biomarkers and modelling of tumour growth towards a personalised diagnosis, prognosis and therapies follow-up.	De-identified medical images, molecular biomarkers and clinical data and extracted QIBs for paediatric cancers	Clinical partners, molecular scientists: biologists, geneticists and image analysts in the consortium
Valencia Imaging Biobank http://www.ceib.san.gva.es/bimcv	Medical Imaging Databank of the Valencia Region (BIMCV) is an image bank conceived as a knowledge repository designed to promote technological advances in medical imaging and to provide technology coverage services to support R&D projects. Currently, BIMCV has a regional scope including all centres in the Valencia Region. However, this management system that integrates digital medical image will give support to the National Health System (SNS) central node and other international nodes.	De-identified images and associated clinical records	Clinical partners and experts
The Cancer Imaging Archive[3] https://www.cancerimagingarchive.net/	US national coverage. Specific for cancer imaging Successful initiative by the U.S. National Cancer Institute's since 2011.	De-identified cancer images Genomics and clinical data	
Radiomics Imaging Archive https://www.radiomicsimagingarchive.eu	RIA is a repository which stores and hosts a large archive of de-identified medical and preclinical images as well as radiomics features extracted from	De-identified medical and preclinical images as well as radiomics features. Currently holds four	Clinical partners and image analysts

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

	these images accessible for public download. It is the European GDPR compliant counterpart to The Cancer Imaging Archive (TCIA) with the difference that it is not limited to oncology or data format.	collections: Liver images during arterial phase segmented; Murine RT-induced lung fibrosis; HX4 rhabdomyosarcoma rat TH302; and Lung1 curated	
UK Biobank imaging[4] https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/imaging-data	UK Biobank is a large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million UK participants	MR images of the heart, MR images of the body, full body (neck to knee) DXA scans, carotid ultrasound images, OCT imaging	
The German National Cohort[5] https://www.dkfz.de/en/epidemiologie-krebserkrankungen/units/NAKO_Studienzentrum_eng/NAKO_engl.html#:~:text=The%20German%20National%20Cohort%20(%E2%80%9CNAKO,participants%20over%2025%2D30%20years.	The German National Cohort (GNC) has been inviting men and women aged between 20 and 69 to 18 study centers throughout Germany since 2014. The participants are medically examined and questioned about their living conditions. There is a focus on seven major disease groups: cancer, diabetes, and cardiovascular, neurologic and psychiatric, infectious, respiratory and musculoskeletal diseases.	not there yet	
NLM's MedPix database https://medpix.nlm.nih.gov/home	Free open-access online database of medical images, teaching cases, and clinical topics, integrating images and textual metadata including over 12,000 patient case scenarios, 9,000 topics, and nearly 59,000 images.	Medical images, teaching cases, and clinical topics. Not only oncology	
EUSOMA database https://www.eusoma.org/en/data-centre/1-345-1-	Eusoma has created in 2006 a central data warehouse of prospectively collected information (eusomaDB) that includes individual records on primary	170,000 cases from 50 European Breast Centres from 11 countries.	Data provided by European Breast Centres that, within the certification process, have provided data to Eusoma according to the society requirements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

	breast cancer cases diagnosed and treated in the European Breast Centres		
NOT ONCOLOGY			
ADNI[6] http://adni.loni.usc.edu/about/	The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimers disease (AD).	Data on Demographics. Clinical asessments and Cognitive assessments; Illumina SNP Genotyping; MRI Images; PET Images; Biospecimen	ADNI researchers
brainlife.io	Free cloud platform for secure neuroscience data analysis.	Not available yet	
FITBIR https://fitbir.nih.gov/	The Federal Interagency Traumatic Brain Injury Research (FITBIR) informatics system: MRI, PET, Contrast, and other data on a range of TBI conditions	The FITBIR Informatics System includes a full range of data from TBI (Traumatic Brain Injury) patients and controls enrolled in studies funded by the DoD and NIH. This data set includes demographics, outcome assessments, imaging and biomarkers.	Data from studies funded by the DoD and NIH
SICAS Medical Image Repository https://www.smir.ch/	SICAS offers a unique combination of competence in acquiring and storing medical images, in processing and visualising data for research and applications in medicine.	2 datasets: -The HEAR-EU Multiscale Imaging and Modelling Dataset of the Human Inner Ear -Full Body CT	
Johns Hopkins Medical Institute's DTI collection https://cmrm.med.jhmi.edu/	A DTI (diffusion tensor imaging) database for normal human brains. The atlas and database will be accessible through this website, which will facilitate the development of various	3 D imaging of developing mouse brains; High-resolution anatomical scans of mouse brains; Blood vessels; Kidney; Imaging	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

	data analysis and visualization tools and contribute to the understanding of the white matter anatomy.	and segmentation of cervical disc herniation. This is a DTI database acquired under Human Brain Project and National Research Resource Center grant. It contains raw and processed DTI data of normal population. Currently we have 2.5 mm isotropic resolution images and 2.2 mm isotropic resolution images. Only 2.5 mm data are available from this site. This database is open to public once the user is registered. The purpose is to facilitate researches in DTI data processing and analysis. The data may also be used to study specific biological interests or as control data. Basic imaging parameters can be also downloaded. Detailed parameters are available upon request.	
Stanford AI in Medicine Database https://aimi.stanford.edu/research/public-datasets	The Stanford Center for Artificial Intelligence in Medicine and Imaging (AIMI Center) was established in 2018 with the primary mission to solve clinically important problems in medicine using AI. Drawing on Stanford's interdisciplinary expertise in clinical medical imaging,	Chest X-rays; Dynamic Cardiac Ultrasound; Lower Extremity RAdiographs; Knee MRI's; MSK Xrays	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

	bioinformatics, statistics, electrical engineering, and computer science, the AIMI Center supports the development, evaluation and dissemination of new AI methods applied across the medical imaging life cycle.		
The Osteoarthritis Initiative https://nda.nih.gov/oai	The Osteoarthritis Initiative (OAI) is a multi-center, ten-year observational study of men and women, sponsored by the National Institutes of Health (part of the Department of Health and Human Services). The goals of the OAI are to provide resources to enable a better understanding of prevention and treatment of knee osteoarthritis, one of the most common causes of disability in adults.	Natural history database for osteoarthritis that will include clinical evaluation data, radiological (x-ray and magnetic OMB Control Number: 0925-0667 Expiration Date: 11/30/2020 December 14, 2017 resonance) images, and a biospecimen repository from 4796 men and women ages 45-79 enrolled between February 2004 and May 2006. Four 3.0 Tesla MRI scanners, one at each clinical center, are dedicated to imaging the knees of OAI participants annually over four years	
AI for COVID imaging archive[7] https://aiforcovid.radiomica.it/	The AIforCOVID imaging archive hosts a large archive of medical images of Italian COVID-19 patients. This project was promoted by CDI Centro Diagnostico Italiano (Milan) in partnership with Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (Milan), Fondazione IRCCS Policlinico San Matteo (Pavia), Azienda	Chest X-Ray of COVID-19 positive patients (DICOM + clinical data)	Radiologists, physicists, engineers

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

	ospedaliero-universitaria Careggi (Florence), ASST Santi Paolo e Carlo (Milan), ASST Fatebenefratelli-Sacco (Milan), ASST Ospedale San Gerardo (Monza), Ospedale Casa Sollievo della Sofferenza (San Giovanni Rotondo), Università Campus Bio-Medico (Roma), and Istituto Italiano di Tecnologia (Genova).		
Dementias Platform UK Data Portal[8] https://portal.dementiasplatform.uk/	anonymized scans of the brain from the UK with virtual environment to run own analysis, dataset accessible upon request	MRI, PET, MEG, 3370929 images	
IMAGENhttps://imagen-europe.com/	brain MRI to study mental health disorders	MRI of the brain	
Neuro C-BIG repository[9] https://www.mcgill.ca/neuro/open-science/c-big-repository	Canadian repository with clinical data of neurological disorders	MRI	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

3.2 FAIR

Name	**Findable** -Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data? -Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset? -Did you provide rich and detailed additional documentation?	**Accessible** - Is the metadata publicly accessible even if the data is no longer available? -Does your dataset contain personal data? -Which of the usage licenses provided by EASY did you choose in order to comply with the access rights attached or the data?	**Interoperable** - Are the data in your dataset stored in preferred formats? - Do you link to other (meta)data and is this (meta)data online resolvable? -Did you provide contextual information about your dataset?	**Reusable** -What kind of information did you provide about the provenance of your data? -Which usage license provided by EASY did choose for your dataset? -Does your (meta)data meet domain standards?
PRIMAGE[2]	Metadata available: Data are described in a so called e-form. Imaging data will be included following a DICOM-MIABIS integration model	Restricted to consortium	The DICOM standard is used to ensure imaging interoperability between different systems in the project. For clinical data management, EUPID integration, and communications between services involved in the project, the HTTP REST standard is used. In addition, standard terminologies like RADLEX or MESH are utilized to annotate relevant data from DICOM headers and clinical variables. Lastly, to improve interoperability, the results will	Clinical partners and experts

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

			also be provided in DICOM SR format which allows systems to perform advanced searches and data mining on DICOM data.	
Valencia Imaging Biobank	Metadata available	Open to external users by use of form and after approval.	DICOM data and metadata are available after anonymization	Control by the research group at La Fe University and Polytechnic hospitalet
Cancer Imaging Archive[3]	Metadata available, standardisation of the modalities to find more easily the datasets (search bar)	Open and limited access collections	Dataset stored in preferred format: csv, dicom, nifti; the metadata is also available on restricted collections; the description of the dataset and the link to the study are provided	versions and usage licence are available, not data processing
Radiomics Imaging Archive	Metadata available	Open and restricted collections	Dataset stored in prefered format: csv, dicom, nifti; the metadata is also available on restricted collections	versions and usage licence are available, not data processing
UK Biobank imaging[4]	Metadata available	Access the database via the Access Management System (AMS) after submitting application, need to pay a fee to access the data	- origin of the data specified, no citations necessary, detailed description (protocols of acquisition, imaging data files generated...)	No information on the format of the images
The German National Cohort[5]	No information	No information	No information	No information on the format of the images
NLM's MedPix database	Metadata available but difficult to select the data as it is per diagnosis only	The data can only be displayed and not downloaded	Origin of the data, extensive description, reference to use	The format of the data is not specified, the follow-up doesn't seems up to date
EUSOMA database		Registered users	No information	No information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

NOT ONCOLOGY				
ADNI[6]	No information on the metadata	A valid user account and an approved ADNI Data Use Application are required to access the ADNI images and data.	No information	Data use agreement available, no information on the metadata
FITBIR		Restricted access. Data access request from is reviewed by a Data Access Quality Committee	No access	Allows versioning of the studies with "metastudy"
SICAS	All the metadata is available	Restrictions according to the user: external viewer,registered user and owner of the repository	Extensive description of the cases	Nifti and dicom format
Johns Hopkins Medical Institute's DTI collection		For registered users only	No access	The dataset contains raw and processed data
Stanford AI in Medicine Database				
The OAI		After registering and obtaining a password, people may browse, download, and use the data for human subjects research for scientific and educational purposes.	There is no access to the image names.	The datasets are described
AI for COVID imaging archive[7]	All the metadata is available	Access on demand	Extensive description of the cases	The dataset contains raw data
Dementias Platform UK Data Portal[8]	No information available without registration	Access on demand	No information available without registration	No information available without registration
IMAGEN	No information available without registration	Access on demand	No information available without registration	No information available without registration

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

Neuro C-BIG repository[9]	No information available without registration	Access on demand	No information available without registration	No information available without registration

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

3.3 Extra databases

Repository	Access	Imaging	Link
Taiwan biobank	Controlled access	Bone density	https://www.twbiobank.org.tw/test_en/index.php
Oxford biobank	Controlled access	Dual-energy X-ray Absorptiometry (DEXA) scan (for fat)	https://www.oxfordbiobank.org.uk/
Qatar biobank	Controlled access	3D carotid scan, iDexa (fat)	https://www.qatarbiobank.org.qa/
Databiobank OncoLifeS	Controlled access	CT, MRI	www.oncolifes.nl Trial
Biomarkpad	Controlled access	MRI, PET	http://jpnd.arone.com/
BioHeart	Not found	CT coronary angiography (CTCA)	Kott KA, et al. doi: 10.1136/bmjopen-2018-028649
AAA Express Biobank Study	Controlled access	CT and MRI	https://clinicaltrials.gov/ct2/results?cond=NCT03320408&term=&cntry=&state=&city=&dist=
BioBank Maastricht UMC: The Maastricht study	Controlled access	MRI, ultrasound	https://www.demaastrichtstudie.nl/research
Joinville stroke biobank	Not found	CT, MRI, angiography, ultrasound	Ferreira LE, et al. doi: 10.1590/0004-282X20170157
Cimbi biobank	Controlled access	MRI, fMRI	https://www.cimbi.dk/index.php
The Polish MOBIT study (Molecular Biomarkers for Individualized Therapy)	Not found	MRI, PET	Niklinski J, et al. doi: 10.1016/j.advms.2017.05.002
Osaka University Twin Registry	Controlled access	MRI, fMRI, PET, ultrasound	Honda C, et al. doi: 10.1017/thg.2019.70
SIREN biobank	Controlled access	CT, MRI	https://catalog.h3africa.org/
PopGen biobank	Controlled access	MRI	https://www.epidemiologie.uni-kiel.de/biobanking/biobank-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

			popgen
The Guangzhou Biobank Cohort Study	not found	Ultrasound, X-Rays	https://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/Guangzhou/index.aspx
Healthy Brain Network biobank	Controlled access	MRI	https://childmind.org/blog/healthy-brain-network-biobank-bringing-big-data-child-mental-health/
Autism brain imaging data exchange	Controlled access	MRI	http://fcon_1000.projects.nitrc.org/indi/abide/
Aragon Workers Health Study (AWHS)	Controlled access	Ultrasound, Calcium score	https://ec.europa.eu/eip/ageing/commitments-tracker/a3/aragon-workers-health-study-awhs_en
The French glioblastoma biobank	Controlled access	MRI	https://www.linkedin.com/in/french-glioblastoma-biobank-808508153/
Erasmus Rotterdam Gezondheid Onderzoek: Rotterdam study	Controlled access	X-rays, Dual-Energy X-ray Absorptiometry	http://www.epib.nl/research/ergo.htm
Hepatitis Biobank at Southwest Hospital	Not found	CT, MRI, ultrasound	Wang, H., et al. doi: 10.1038/srep38180
Amsterdam Dementia Cohort	Not found	MRI, PET	https://www.neurodegenerationresearch.eu/cohort/amsterdam-dementia-cohort/
UMCG Research Data and Biobanking Team: ROBINSCA trial	Not found	CT	https://cordis.europa.eu/project/id/294604/reporting/de
ImaLife	Controlled access	CT	https://www.lifelines.nl/researcher/cohort-and-biobank/news-2/2-3
Health and Prevention Enhancement	Controlled access	CT, MRI, X-Rays	http://en-healthcare.snuh.org/HPEACEstudy
Brain Genomics Superstruct Project	Public access	MRI	https://dataverse.harvard.edu/dataverse/GSP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

Swedish CArdioPulmonarybioImage Study	Controlled access	CT, MRI	https://snd.gu.se/en/catalogue/study/ext0054
German Diabetes Study	Not found	MRI	https://www.dzd-ev.de/en/research/multicenter-studies/gds/index.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.

4. Discussion and conclusion

We have created an extensive (but not exhaustive) map of existing health data repositories known to the CHAIMELEON consortium, and analysed it for the purpose of incorporation into a wider umbrella repository. The severely fragmented nature of these repositories is based on underlying legal and ethical issues associated with health data, which in many European countries is considered to be the property of the patient. Information about accessibility and interoperability is not always available, but it is clear that in the majority of cases these repositories were designed and act as stand-alone entities, with integration into a larger system problematic at least. Nonetheless, the necessity of a FAIR, GDPR compliant, European repository for AI ready data still exists, and we have come up with a few lessons learnt from this exercise.

4.1 Lesson learnt

The lessons learned can be used as input of future deliverables:

- **D3.3- Interim platform design**
- **D3.4 Final platform design**

Based on our assessment of the existing repositories we came up with the following points that need to be taken into account in setting up the CHAIMELEON repository:

- Most studies will need to be hosted locally at the center providing the data, with rigid access control, so a distributed storage solution with a central search engine will be necessary. Furthermore, due to legal and financial reasons, different centers will be using different storage solutions, so the repository design should focus more on how to collate, catalogue, and present the data within each silo. A simple solution would be the generation of manifests upon data ingestion (in the form of an uploader tool) that get updated centrally. This is also the solution we finally chose for RIA.

- Datasets are more interesting as whole studies, since individual images are not of particular importance to AI application. The platform design should reflect this and be optimized for searching and fast download of entire collections, not of individual patients. Datasets need to be well described and have DOIs.

- A clear separation between private and open source datasets needs to be made, and access control needs to be one major aspect of the repository to ensure that only those who are allowed to see / use data have access.

- Data science has progressed into many directions, with many disparate methodologies and using many separate data formats. Restricting CHAIMELEON to only one format (e.g. DICOM) would severely reduce the amount of datasets that can be stored.

- Most datasets require extensive data curation of download. It would aid the users if datasets included in the CHAIMELEON repository had already been curated. Hence, the CHAIMELEON repositories should include at least a couple of “ready to use”data sets. For the datasets that still require some form of curation the level of curation could be described.

- Many AI studies require validation in another dataset. It would aid researchers if, if possible, it would be clear which data set could be used for validation if they choose a certain dataset. In other words, the CHAIMELEON repository could include “pairs”of datasets, one for training and one for validation.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 952172.

4.2 Conclusion

The list provided and the conclusions drawn from the dataset and experience are by no means exhaustive but provide enough evidence to make tentative suggestions. Due to the private nature of health records, most repositories severely restrict access to most databases, and as such require the use of a front-end provided by the repository to access the data. Furthermore, at the back-end there is a plethora of systems being used for storage and indexing, from the PACS-like XNAT system to those developed in-house. As such, the inclusion of these databases into a larger network would require either the formation of a new consortium including the repositories of interest, or a massively complex infrastructure that deals with each repository as a separate project. We believe that the way forward is a de-centralized repository with a central index and search portal.

References

[1]http://www.aylward.org/notes/open-access-medical-image-repositories

[2] Scapicchio, C., Gabelloni, M., Forte, S.M. et al. (2021) DICOM-MIABIS integration model for biobanks: a use case of the EU PRIMAGE project. Eur Radiol Exp 5, 20. https://doi.org/10.1186/s41747-021-00214-4

[3] Coppola, L., Cianflone, A., Grimaldi, A. M., Incoronato, M., Bevilacqua, P., Messina, F., ... & Salvatore, M. (2019). Biobanking in health care: evolution and future directions. Journal of translational medicine, 17(1), 1-18.

[4] Sudlow, C., Gallacher, J., Allen, N., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. Plos med, 12(3), e1001779.

[5] Schipf, S., Schöne, G., Schmidt, B., et al. (2020). The baseline assessment of the German National Cohort (NAKO Gesundheitsstudie): participation in the examination modules, quality assurance, and the use of secondary data. https://doi. org/10.1007/s00103-020-03093-z.

[6] Petersen, R. C., Aisen, P. S., Beckett, L. A., et al. (2010). Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology, 74(3), 201-209.

[7]Soda, P., D'Amico, N. C., Tessadori, J., et al. (2020). AIforCOVID: predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study.

[8] Bauermeister, S., Orton, C., Thompson, S., et al. (2020). The dementias platform UK (DPUK) data portal. European journal of epidemiology, 35(6), 601-611.

[9] Das, S., Glatard, T., Rogers, C., et al. (2017). Cyberinfrastructure for open science at the Montreal Neurological Institute. Frontiers in neuroinformatics, 10, 53.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952172.