IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Bridging gaps between images and data: a systematic update on imaging biobanks

Michela Gabelloni[1] · Lorenzo Faggioni[1] · Rita Borgheresi[1] · Giuliana Restante[1] · Jorge Shortrede[1] · Lorenzo Tumminello[1] · Camilla Scapicchio[1] · Francesca Coppola[2,3] · Dania Cioni[1] · Ignacio Gómez-Rico[4] · Luis Martí-Bonmatí[4,5] · Emanuele Neri[1,3]

Abstract

Background and objective The systematic collection of medical images combined with imaging biomarkers and patient non-imaging data is the core concept of imaging biobanks, a key element for fuelling the development of modern precision medicine. Our purpose is to review the existing image repositories fulfilling the criteria for imaging biobanks.

Methods Pubmed, Scopus and Web of Science were searched for articles published in English from January 2010 to July 2021 using a combination of the terms: "imaging" AND "biobanks" and "imaging" AND "repository". Moreover, the Community Research and Development Information Service (CORDIS) database (https://cordis.europa.eu/projects) was searched using the terms: "imaging" AND "biobanks", also including collections, projects, project deliverables, project publications and programmes.

Results Of 9272 articles retrieved, only 54 related to biobanks containing imaging data were finally selected, of which 33 were disease-oriented (61.1%) and 21 population-based (38.9%). Most imaging biobanks were European (26/54, 48.1%), followed by American biobanks (20/54, 37.0%). Among disease-oriented biobanks, the majority were focused on neuro-degenerative (9/33, 27.3%) and oncological diseases (9/33, 27.3%). The number of patients enrolled ranged from 240 to 3,370,929, and the imaging modality most frequently involved was MRI (40/54, 74.1%), followed by CT (20/54, 37.0%), PET (13/54, 24.1%), and ultrasound (12/54, 22.2%). Most biobanks (38/54, 70.4%) were accessible under request.

Conclusions Imaging biobanks can serve as a platform for collecting, sharing and analysing medical images integrated with imaging biomarkers, biological and clinical data. A relatively small proportion of current biobanks also contain images and can thus be classified as imaging biobanks.

Key Points
- *Imaging biobanks are a powerful tool for large-scale collection and processing of medical images integrated with imaging biomarkers and patient non-imaging data.*
- *Most imaging biobanks retrieved were European, disease-oriented and accessible under user request.*
- *While many biobanks have been developed so far, only a relatively small proportion of them are imaging biobanks.*

Keywords Diagnostic imaging · Biomarkers · Artificial intelligence

---

Michela Gabelloni and Lorenzo Faggioni are co-first authors.

Luis Martí-Bonmatí and Emanuele Neri are co-last authors.

✉ Lorenzo Faggioni
lfaggioni@sirm.org

1   Academic Radiology, Department of Translational Research, University of Pisa, via Roma, 67, 56126 Pisa, Italy

2   Department of Radiology, IRCCS Azienda Ospedaliero-Universitaria Di Bologna, 40138 Bologna, Italy

3   Italian Society of Medical and Interventional Radiology, SIRM Foundation, via della Signora, 2, 20122 Milan, Italy

4   Biomedical Imaging Research Group (GIBI230-PREBI) at La Fe Health Research Institute, 46026 Valencia, Spain

5   Medical Imaging Department, Hospital Universitario y Politécnico La Fe, 46026 Valencia, Spain

Abbreviations

AI	Artificial intelligence
BBMRI-ERIC	Biobanking and BioMolecular resources Research Infrastructure — European Research Infrastructure Consortium
CARDIATEAM	CARdiomyopathy in type 2 DIAbetes mellitus
COPD	Chronic obstructive pulmonary disease
CORDIS	Community Research and Development Information Service
DIPG	Diffuse intrinsic pontine glioma
DMG	Diffuse midline glioma
DXA	Dual-energy X-ray absorptiometry
ESR	European Society of Radiology
FAIR	Findability, Accessibility, Interoperability, and Reusability
GDPR	General Data Protection Regulation
HPCI	High-performance computing infrastructure
MESA	Multi-Ethnic Study of Atherosclerosis
MIABIS	Minimum Information About BIobank data Sharing
Mobit	Molecular Biomarkers for Individualized Therapy
OMOP CDM	Observational Medical Outcomes Partnership Common Data Model
PRIMAGE	PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Introduction

According to the statutes of the Biobanking and BioMolecular resources Research Infrastructure — European Research Infrastructure Consortium (BBMRI-ERIC), biobanks are defined as "collection, repositories and distribution centres of all types of human biological samples, such as blood, tissues, cells or DNA and/or related data such as associated clinical and research data, as well as biomolecular resources, including model- and micro-organisms that might contribute to the understanding of the physiology and diseases of humans" [1].

The terms "biobank" and "biorepository" are often used interchangeably. According to the National Institutes of Health (NIH), biorepositories are "a place, room, or container where biospecimens/tissue samples are stored", where the term "biospecimen" refers to materials taken from the human body (such as tissues, blood, plasma and urine), commonly accompanied by information about the patient from whom the biospecimen was obtained [2]. More specifically, while biorepositories may be collections of biological material from any living organism, biobanks usually refer to collections of human biological material [3, 4]. In order to gather such a plenty of data, it has been necessary to develop large electronic databases [5]. Based on the Declaration of Taipei published by the World Medical Association, a health database can be defined as "a system for collecting, organising and storing health information" [6]. Imaging data were initially either not collected in biobanks or minimally represented. For example, one of the first cohort studies involving the collection of MRI data was the Multi-Ethnic Study of Atherosclerosis (MESA), which included less than 5000 patients with the goal to investigate subclinical cardiovascular diseases in the general population [7–9]. Only in the last few years, more imaging data have begun to be collected in biobanks, which are often intended as a collection of biological "images", not necessarily related to radiology, but consisting of digital images of pathology specimens [9]. To address the concept of imaging biobanks in the field of precision medicine, the European Society of Radiology (ESR) Research Committee established an Imaging Biobanks Working Group in 2014, which defined imaging biobanks as "organised databases of medical images and associated imaging biomarkers (radiology and beyond), shared among multiple researchers, linked to other biorepositories", and suggested that "biobanks (which focus only on the collection of genotype data) should simultaneously come with a system to collect related clinical or phenotype data" [10]. The primary aim of imaging biobanks is to guarantee the long-term storage and retrieval of secured medical images and associated metadata for research and validation purposes, whereas the secondary goal is to connect imaging and tissue biobanks, thus providing a deep association between phenotype and genotype expressions [11, 12].

The BBMRI differentiates between two types of research biobanks, i.e. population-based and disease-oriented biobanks. Population-based biobanks collect data from the general population and are focused on identifying risk factors for disease development [such as the Rotterdam Study (https://populationimaging.eu/image-data/)], whereas disease-oriented biobanks aim to investigate the pathogenesis of human diseases and are generally focused on specific diseases (mostly cancer, such as the Primage, CHAIMELEON, Pro-Cancer-I and EuCanImage H2020 projects [13–16]).

Modern biobanks are more than a collection of samples and data, but a complex infrastructure in which biomedical images and clinical data extracted from medical health records are organised, curated through standard procedures and securely protected from unauthorised external access [17–21]. All these aspects must coexist in compliance with current European legislation on data protection (such as the

GDPR), in addition to the national laws of the countries involved in the construction and use of biobanks. Biobanks need to generate data governance models that protect the information provided by the subjects enrolled, while simultaneously enabling and promoting its use for biomedical research purposes with open controlled access to data sets [22]. To overcome fragmentation, the most prominent European biobanks have started collaborations within BBMRI, whose mission is to secure access to biological resources and data required for health-related research in a sustainable manner.

In 2015, the ESR Working Group on Imaging Biobanks evaluated imaging biobanks within the ESR community by administering a survey to its members. The survey identified 27 potential repositories of images and associated biomarkers that could be in line with the definition of imaging biobanks [10]. However, no specific details regarding each biobank could be retrieved at the time of the survey, and the latter has received no updates so far. Although few articles overview imaging biobanks [5], they are focused on specific types of biobanks, such as population-based [23] or oncological biobanks [11]. To the best of our knowledge, no articles have been published to date which offer a systematic review of existing imaging biobanks. Our purpose is to explore the current status of imaging biobanks.

Methods

Literature search strategy

A systematic literature search was carried out to identify imaging biobanks. To this purpose, we searched the PubMed (https://pubmed.ncbi.nlm.nih.gov), Scopus (https://www.scopus.com) and Web of Science databases (https://www.webofscience.com) using the combination of the following search terms: "imaging" AND "biobanks", and in a separate search, "imaging" and "repository". The search period ranged from January 2010 to July 2021, and the search was carried out in July 2021. Only articles published in English were selected.

Moreover, the Community Research and Development Information Service (CORDIS) database (https://cordis.europa.eu/projects) was searched using the combination of the search terms: "imaging" AND "biobanks", with the search encompassing collections, projects, project deliverables, project publications and programmes.

Study selection

In an attempt to improve the quality of our inclusion criteria, our analysis was performed following a four-step flow diagram based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [24].

Results

All retrieved publications (totalling 3139 and 6133 articles regarding imaging biobanks and imaging repositories, respectively) were separately uploaded to EndNote™, and all duplicate records were removed (1157 and 2107 articles regarding imaging biobanks and imaging repositories, respectively).

Two reviewers (M.G., R.B.) manually screened all included articles by title and abstract, and if potentially eligible, their full text was reviewed. The reviewers excluded (a) entries including both terms (imaging AND biobanks or imaging AND repository) but not necessarily related to each other in the text, (b) entries including imaging but unrelated to diagnostic imaging, (c) reviews and conference papers, (d) entries including the same imaging biobank and (e) entries including biobanks with data derived from imaging, but without image collection (Fig. 1). Finally, 42 articles with a reference to an imaging biobank and 46 articles with a reference to an imaging repository (of which 34 referring to the same biobank) were selected, resulting in a total of 54 imaging biobanks.

To obtain more detailed information on the storage and use of images by the biobanks selected, the biobank website (if available) or the trial or project pages were also accessed.

Of the 54 imaging biobanks retrieved, 21 were population-based (38.9%) [9, 25–44] and 33 were disease-oriented (61.1%) [45–77]. Most imaging biobanks were European (26/54, 48.1%), followed by American (20/54, 37.0%) biobanks, with the remaining ones being based in Asia (5/54, 9.3%), Oceania (2/54, 3.7%) and Africa (1/54, 1.9%). Thirty-eight out of 54 biobanks (70.4%) were accessible under request, 1 (1.8%) supported public access, and for the remaining 15 (27.8%) no information was found (Fig. 2).

The number of patients collected by each biobank ranged from 240 to 3,370,929 (last access date: 14 July 2021) (Table 1). Among disease-oriented biobanks, 9 were focused on neurodegenerative diseases (9/33, 27.3%), with particular regard to Alzheimer disease and other dementias, and 9 on cancer (9/33, 27.3%), of which 3 related to multiple cancers, 2 to lung cancer, 2 to glioma, 1 to diffuse intrinsic pontine glioma (DIPG) and diffuse midline glioma (DMA) and 1 to neuroblastoma and DIPG. The remaining disease-oriented biobanks were focused on the following: other neurological and psychiatric diseases (6/33 (18.2%), including migraine (1), autism (1) and neuropsychiatric disorders (2), traumatic brain injuries (1) and post-traumatic epilepsy (1)), stroke (3/33, 9.1%), cardiovascular diseases and diabetes (3/33, 9.1%) and lung diseases excluding cancer (2/33 (6.1%),
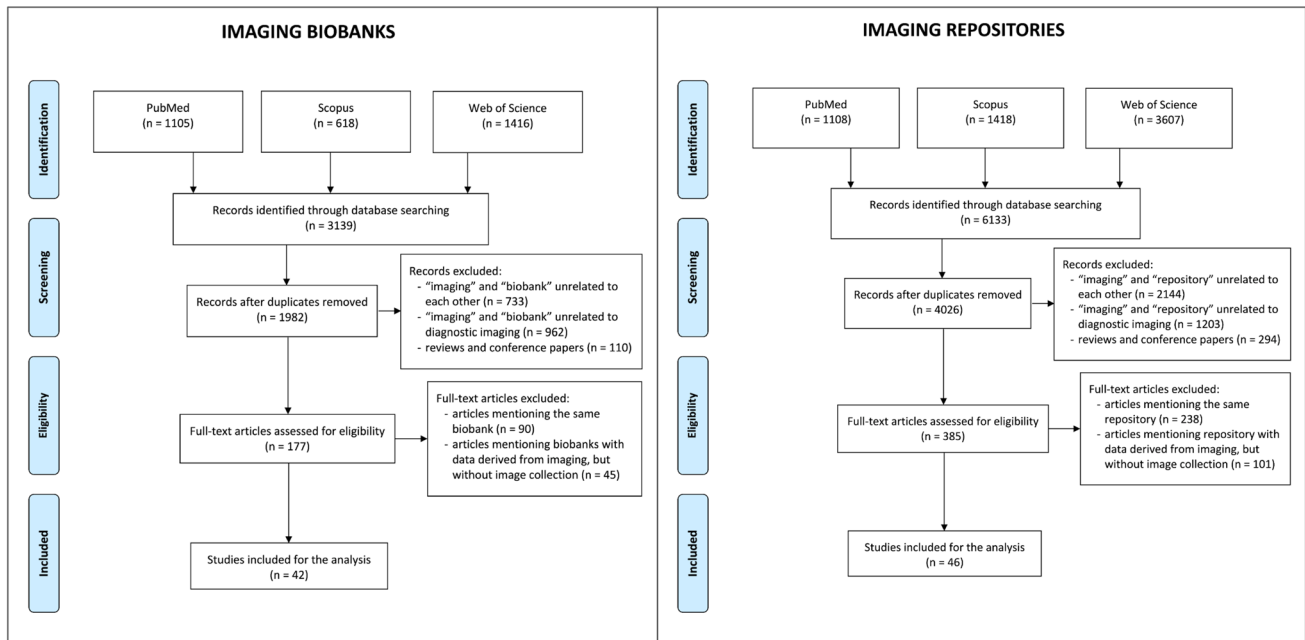
Fig. 1 PRISMA flow chart of literature search for imaging biobanks (left) and imaging repositories (right). Adapted from [24]
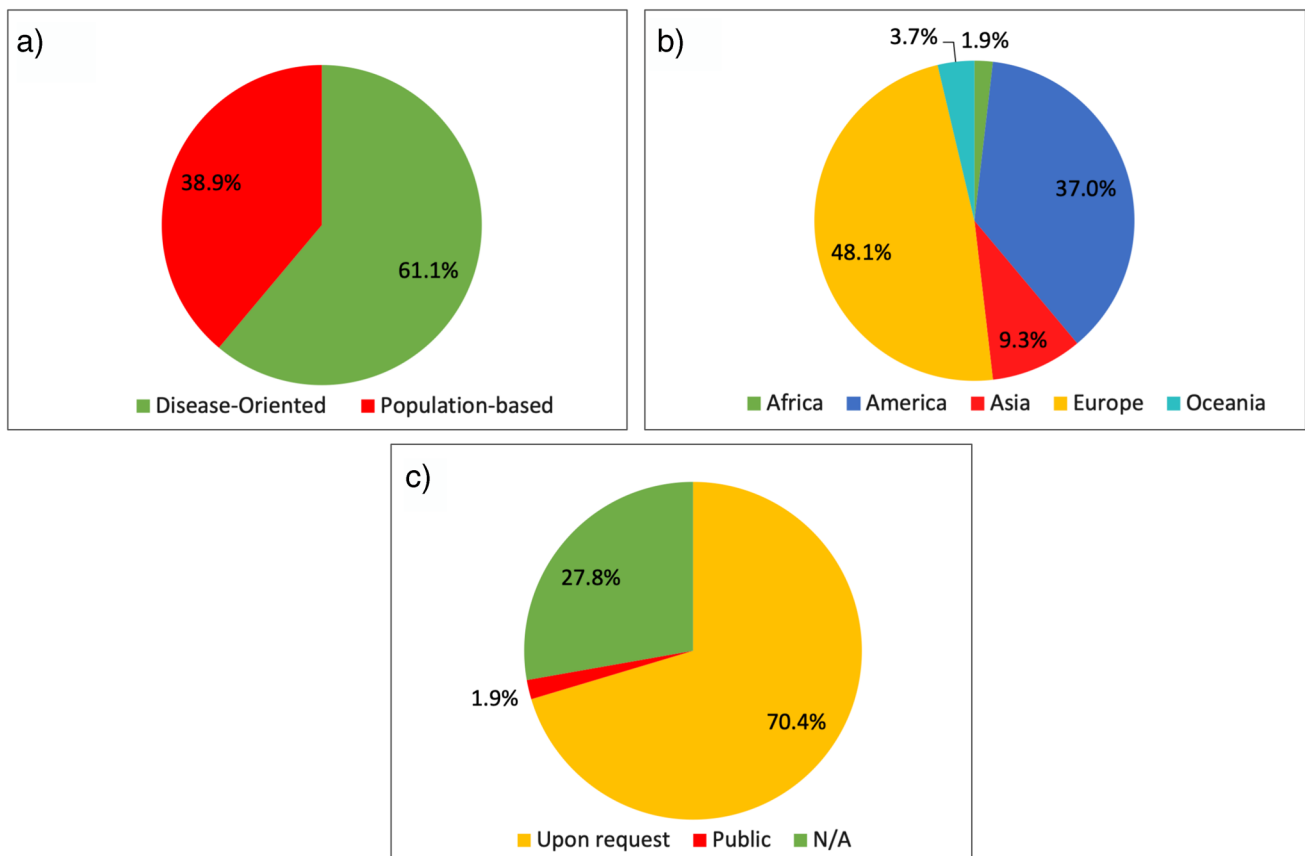

Fig. 2 Pie charts of existing imaging biobanks classified by (**a**) target population (i.e. disease-oriented and population-based), (**b**) geographical distribution and (**c**) type of data access

Table 1 List of existing imaging biobanks classified based on their distinguishing features. COPD = chronic obstructive pulmonary disease, DIPG = diffuse intrinsic pontine glioma, DMG = diffuse midline glioma, DO = disease-oriented, DSA = digital subtraction angiography, HR-pQCT = high-resolution peripheral quantitative computed tomography, PB = population-based

Imaging biobank	Continent	Data access	PB	DO	Imaging modality	No. of patients	Reference	Disease	Tissue(s) of interest	Year
Adolescent Brain Cognitive Development (ABCD)	America	Upon request	X		MRI	11,878	https://abcdstudy.org/about/	Brain development	Brain	2016
Akershus Cardiac Examination (ACE) 1950	Europe	N/A	X		Echocardiography, ultrasound	3706	https://ace1950.no/english	Cardio- and cerebrovascular	Cardiovascular system	2012
Alzheimer's Disease Neuroimaging Initiative (ADNI)	America	Upon request		X	MRI, PET	2000+	http://adni.loni.usc.edu/	Alzheimer	Brain	2004
American Registry for Migraine Research (ARMR)	America	Upon request		X	CT, MRI, PET	996	https://americanmigrainefoundation.org/resource-library/american-registry-for-migraine-research/	Migraine	Brain	2016
Amsterdam Dementia Cohort	Europe	N/A		X	MRI, PET	5000+	https://www.neurodegenerationresearch.eu/cohort/amsterdam-dementia-cohort/	Neurodegenerative (dementia)	Brain	2004
Aragon Workers Health Study (AWHS)	Europe	Upon request	X		CT, ultrasound	5400	https://ec.europa.eu/eip/ageing/commitments-tracker/a3/aragon-workers-health-study-awhs_en	Overweight and obesity, diabetes and other cardiovascular risk factors	cardiovascular system	2009
Australian Imaging, Biomarkers and Lifestyle (AIBL) Study	Oceania	Upon request		X	MRI, PET	1000+	https://aibl.csiro.au/	Alzheimer	Brain	2006
Autism Brain Imaging Data Exchange (ABIDE)	America	Upon request		X	MRI	1112	http://fcon_1000.projects.nitrc.org/indi/abide/	Autism	Brain	2012
BioHEART-CT	Oceania	N/A		X	CT coronary angiography	5000	[47]	Cardiovascular	Cardiovascular system	2015
Biomarkpad	Europe	Upon request		X	MRI, PET	419	http://jpnd.arone.com/	Neurodegenerative	Brain	2013
Brain Genomics Superstruct Project (GSP)	America	Upon request	X		MRI	1570	https://dataverse.harvard.edu/dataverse/GSP	Neurobehavioural	Brain	2008
Canadian Alliance for Healthy Hearts and Minds (CAHHM)	America	N/A	X		MRI	9700	https://www.phri.ca/research/cahhm/	Cardio- and cerebrovascular	Body fat, brain, cardiovascular system	2013
CENTER-TBI	Europe	Upon request		X	MRI	5400	https://www.center-tbi.eu/	Traumatic brain injuries	Brain	2014
Cimbi Biobank	Europe	Upon request		X	MRI, PET	800	https://www.cimbi.dk/index.php	Neuropsychiatric disorders	Brain	2008
COSYCONET	Europe	N/A		X	CT	2741	http://www.asconet.net/html/cosyconet	COPD	Lung	2010

Table 1 (continued)

Imaging biobank	Continent	Data access	PB	DO	Imaging modality	No. of patients	Reference	Disease	Tissue(s) of interest	Year
Dementias Platform UK Data Portal (DPUK)	Europe	Upon request		X	MRI, PET	3,370,929	https://portal.dementiasplatform.uk/	Neurodegenerative (dementia)	Brain	2017
DescribePSP and ProPSP	Europe	N/A		X	MRI	354+269	https://www.dzne.de/en/research/studies/clinical-studies/describe/describe-psp/	Progressive supranuclear palsy	Brain	2015
Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx)	America	Upon request		X	CT, MRI	300	https://epibios.loni.usc.edu/	Post-traumatic epilepsy	Brain	2016
French Glioblastoma Biobank	Europe	Upon request		X	MRI	1087	https://fr.linkedin.com/in/french-glioblastoma-biobank	Glioblastoma	Brain	2013
Generation R Study	Europe	N/A	X		MRI, ultrasound	9778	https://generationr.nl/researchers/	N/A	Body fat, brain, cardiovascular, hip, kidney, lung	2002
German Diabetes Study (GDS)	Europe	N/A		X	MRI	698	https://www.dzd-ev.de/en/research/multicenter-studies/gds/index.html	Diabetes	Body fat	2009
German National Cohort (NAKO Gesundheitsstudie)	Europe	Upon request		X	MRI	30,000	https://www.dkfz.de/en/epidemiologie-krebserkrankungen/units/NAKO_Studienzentrum_eng/NAKO_engl.html	Cancer, diabetes, cardiovascular, neurologic, psychiatric, infectious, respiratory and musculoskeletal diseases	Whole body	2014
Guangzhou Biobank Cohort Study	Asia	N/A	X		Ultrasound, X-ray	30,519	https://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/Guangzhou/index.aspx	Chronic diseases	Carotid arteries, chest	2003
Health and Prevention Enhancement (H-Peace) Study	Asia	Upon request	X		CT, MRI, X-ray	91,336	http://en-healthcare.snuh.org/HPEAC Estudy	Non-communicable diseases (such as malignancies and metabolic diseases)	brain, cardiovascular system	2003
Healthy Brain Network Biobank	America	Upon request			MRI	10,000	https://childmind.org/blog/healthy-brain-network-biobank-bringing-big-data-child-mental-health/	Psychiatric disorders	Brain	2015
Huashan Glioma Tissue Bank (GTB)	Asia	N/A		X	MRI	1326+	[64]	Glioma	Brain	2009

Table 1 (continued)

Imaging biobank	Continent	Data access	PB	DO	Imaging modality	No. of patients	Reference	Disease	Tissue(s) of interest	Year
Human Connectome Project	America	Upon request	X		MRI	1200	https://www.humanconnectome.org/study/hcp-young-adult	N/A	Brain	2010
IMAGEN	Europe	Upon request	X		MRI	2000	https://imagen-europe.com/	Neurodegenerative	Brain	2010
ImaLife	Europe	Upon request	X		CT	12,000	https://www.lifelines.nl/researcher/cohort-and-biobank/news-2/2-3	Cardiovascular disease, COPD, lung cancer	Cardiovascular system, lung	2017
International DIPG/DMG Registry (IDIPGR)	America	Upon request		X	CT, MRI	448	https://www.dipgregistry.org/	DIPG, DMG	Brain	2012
Joinville Stroke Biobank	America	N/A		X	CT, Doppler ultrasound, DSA, MRI	2688	[49]	stroke	Brain	2010
Maastricht Study	Europe	Upon request		X	DXA, HR-pQCT, MRI, X-ray	10,000	https://www.demaastrichtstudie.nl/research	Type 2 diabetes	Body fat, bone, brain	2010
MRI-GENetics Interface Exploration (MRI-GENIE)	America	Upon request		X	MRI	3301	http://www.resilientbrain.org/mrigenie.html	Stroke	Brain	2014
Nashville Lung Cancer Screening Trial Cohort	America	N/A		X	CT	410	https://clinicaltrials.gov/ct2/show/NCT01475500	Lung cancer	Lung	2011
National Alzheimer's Coordinating Center (NACC)	America	Upon request		X	MRI, PET	26,586+	https://naccdata.org/	Cognitive disorders (such as Alzheimer's disease, vascular dementia, Lewy body dementia, frontotemporal lobar degeneration)	Brain	2005
Neuro C-BIG Repository	America	Upon request		X	MRI, PET	3000	https://www.mcgill.ca/neuro/open-science/c-big-repository	Neurological	Brain	2020
NIH Pediatric MRI Data Repository	America	Upon request	X		MRI	556	http://pediatricmri.nih.gov	Brain development	brain	2001
OncoLifeS	Europe	Upon request		X	CT, MRI	10,000	https://umcgresearch.org/nl/w/oncolifes	Cancer	Gynaecological, head and neck, lung	2014
Osaka University Twin Registry	Asia	Upon request	X		MRI, PET, ultrasound	900	[36]	N/A	Brain, carotid arteries, thyroid	2009
Oxford Biobank	Europe	Upon request	X		DXA	10,000	https://www.oxfordbiobank.org.uk/	Cardiovascular disease, diabetes, obesity	Body fat	2000

Table 1 (continued)

Imaging biobank	Continent	Data access	PB	DO	Imaging modality	No. of patients	Reference	Disease	Tissue(s) of interest	Year
Parelsnoer Institute Biobanks	Europe	Upon request		X	MRI	32,000+	http://parelsnoer.org/page/en/home	Cancer, cardiovascular and other chronic diseases (e.g. Neurodegenerative and inflammatory bowel diseases, renal failure, osteoarthrosis, rheumatoid arthritis)	Whole body	2007
Parkinson's Disease Biomarkers Program (PDBP)	America	Upon request		X	MRI	240	https://pdbp.ninds.nih.gov/	Dementia with Lewy bodies, Parkinson disease	Brain	2012
Polish MOBIT Project	Europe	N/A		X	PET/MRI	N/A	[51]	Lung cancer	lung	2015
PRIMAGE	Europe	Upon request		X	CT, MRI, PET	2500+	https://www.primageproject.eu/project/	Neuroblastoma, DIPG	Brain, whole body	2018
Progression and Early detection of Subclinical Atherosclerosis (PESA)	Europe	N/A	X		CT, MRI, PET ultrasound	4000+	http://www.estudiopesa.org/en/index.php	Cardiovascular	Cardiovascular system	2010
Qatar Biobank	Asia	Upon request	X		DXA, ultrasound	60,000	https://www.qatarbiobank.org.qa/	Chronic diseases	Body fat, carotid arteries	2016
ROBINSCA Trial	Europe	N/A	X		CT	12,950	https://cordis.europa.eu/project/id/294604/reporting/de	Cardiovascular	Cardiovascular system	2012
Rotterdam Study	Europe	Upon request	X		DXA, ultrasound, X-ray	15,000	http://www.epib.nl/research/ergo.htm	Cardiovascular, neurological, locomotor, ophthalmic	Brain, cardiovascular system	1990
SIREN Biobank	Africa	Upon request		X	CT, MRI	3000+	https://catalog.h3africa.org/	Stroke	Brain	2013
Swedish CArdioPulmonarybioImage Study (SCAPIS)	Europe	Upon request	X		CT, MRI, ultrasound	30,000	https://snd.gu.se/en/catalogue/study/ext0054	Cardiovascular disease and COPD	Body fat, cardiovascular system, lung	2012
TB Portals	America	Upon request		X	CT, X-Ray	5339	https://tbportals.niaid.nih.gov/	Tuberculosis	Lung	2013
TCIA	America	Public		X	CT, MRI, PET, ultrasound, X-ray	58,000+	https://www.cancerimagingarchive.net/	Cancer	Whole body	2010
UK Biobank	Europe	Upon request	X		DXA, MRI, ultrasound	10,000	https://www.ukbiobank.ac.uk/	Cardiovascular and neurodegenerative diseases, cancer, diabetes, obesity, etc	Brain, carotid arteries, heart, visceral fat	2014

Table 1 (continued)

Imaging biobank	Continent	Data access	PB	DO	Imaging modality	No. of patients	Reference	Disease	Tissue(s) of interest	Year
Veterans Affairs Precision Oncology Data Repository (VA-PODR)	America	Upon request		X	CT	113,154	https://vpodc.org	Cancer	Whole body	2016

including tuberculosis (1) and chronic obstructive pulmonary disease (COPD) (1)). Only 1 disease-oriented biobank (1/33, 3.0%) was mixed, due to including data about cancers, cardiovascular and chronic diseases.

The imaging modality most frequently involved was MRI (40/54, 74.1%), followed by CT (20/54, 37.0%), PET (13/54, 24.1%) and ultrasound (12/54, 22.2%) (Fig. 3). Two-third of biobanks had been developed in the consecutive years 2009–2010 (11/54, 20.4%) and 2012–2016 (25/54, 46.3%) (Fig. 4).

The following 3 disease-oriented imaging biobanks were found within the CORDIS database: CHAIMELEON, PRIMAGE and CARDIATEAM [13, 14, 78].

Discussion

Existing imaging biobanks are imaging data repositories intended for research, which also contain clinical data and are mainly disease-oriented, with a more frequent focus on oncologic, neurological and cardiovascular diseases. A major role of current imaging biobanks is to provide imaging data for radiomics analysis so as to obtain biomarkers that can be correlated with clinical, genomics and histopathological factors. Although all biobanks included in our study have imaging data stored in them, only some (e.g. Bioheart Study, Cimbi, Health Brain Network, Alzheimer's Disease Neuroimaging Initiative, ROBINSCA, Imagen) have correlated imaging data (e.g. functional MRI) with clinical and laboratory data, with a few planning on analysing quantitative imaging features and parameters to obtain imaging biomarkers.

Among disease-oriented biobanks focused on cancer, OncoLifeS (Oncological Life Study: Living well as a cancer survivor) is a hospital-based biobank with the threefold goal to offer an infrastructure for clinical cancer research, to drive translational research towards more personalised cancer care and to monitor the quality of oncological care outcomes. It also aims to integrate clinical, laboratory, pathological data and image biomarkers in order to predict a range of outcomes, such as progression and treatment response in cancer patients [45].

Several biobanks have been built with the objective to deal with specific tumour types. Among them, PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers) is a funded Horizon 2020 project with a 4-year duration launched in 2018. The main purpose of this project is to create an open cloud-based platform for supporting decision making in the clinical management of two paediatric cancers, i.e. neuroblastoma (the most frequent solid cancer of early childhood) and DIPG (the leading cause of brain tumour-related death in children) [13].

Fig. 3 Diagrams of existing imaging biobanks classified by disease distribution among disease-oriented biobanks (upper) and imaging modality (lower). *Mixed biobank (Parelsnoer Institute Biobanks)
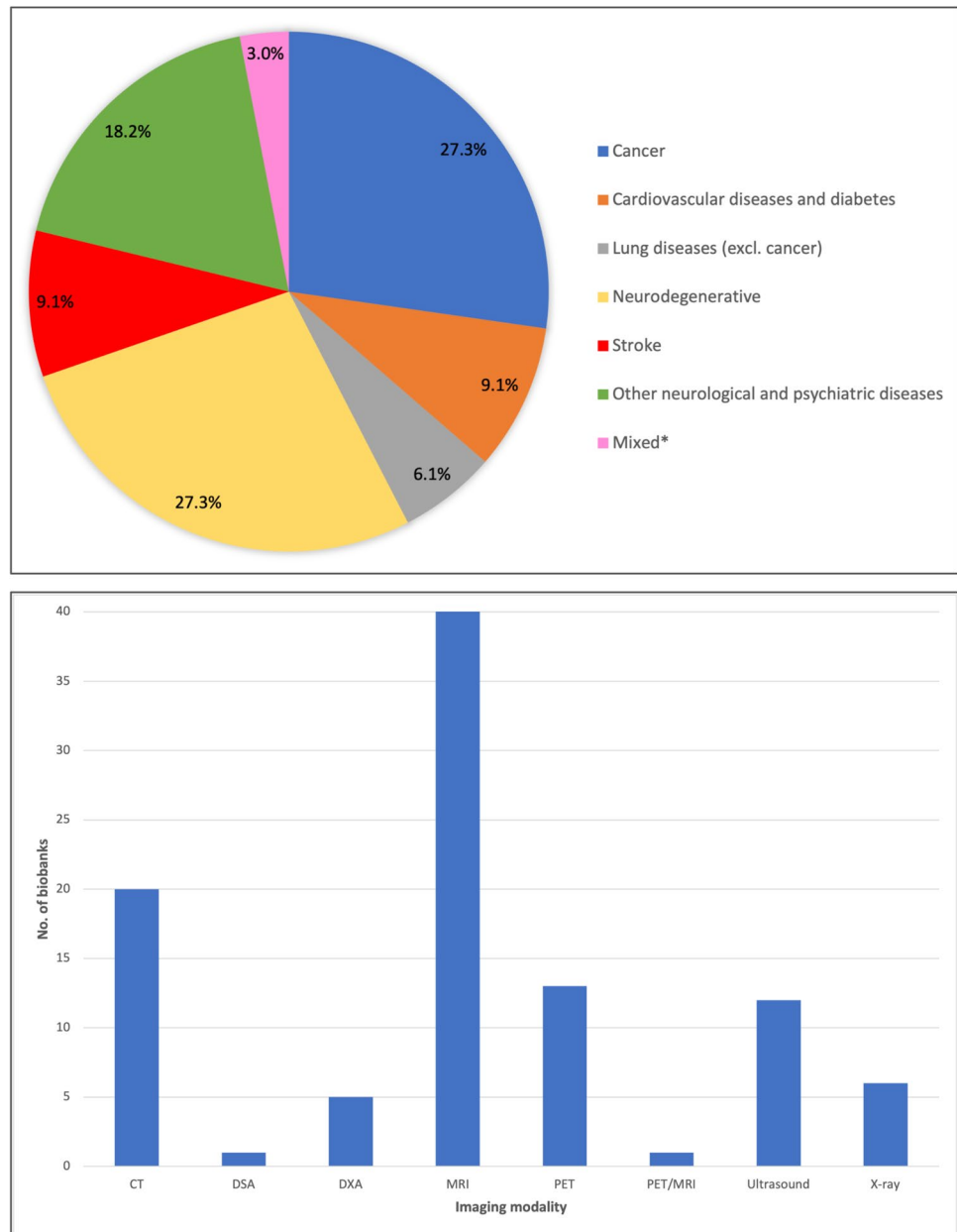
Other tumour-specific biobank projects include the Polish Mobit project, the French Glioblastoma Biobank and the CHAIMELEON project. CHAIMELEON (Accelerating the lab to market transition of AI tools for cancer management) is a funded Horizon 2020 4-year research project started in September 2020 and aimed to develop a structured repository of health images and related clinical and molecular data on the most prevalent cancers in Europe, such as lung, breast, prostate and colorectal cancer [14]. The objectives of the Polish Mobit (Molecular Biomarkers for Individualized Therapy) project include establishing a lung cancer biobank collecting patients' tissue, blood and urine samples, and developing individualised lung cancer diagnostics

which integrate genomics, transcriptomics, metabolomics and PET/MRI radiomics analysis [51]. The French Glioblastoma Biobank is another national biobank founded in 2013 with the goal to collect clinical and imaging data, along with biological samples, for supporting translational research projects and AI technologies for the management of glioblastoma [61].

A disease-oriented biobank project focused on non-oncological diseases is CARDIATEAM (CARdiomyopathy in type 2 DIAbetes mellitus), a 5-year research project started in 2019 and supported by Horizon 2020. The aim of this project is to determine how distinct diabetic cardiomyopathy is from other forms of heart failure, and to assess the
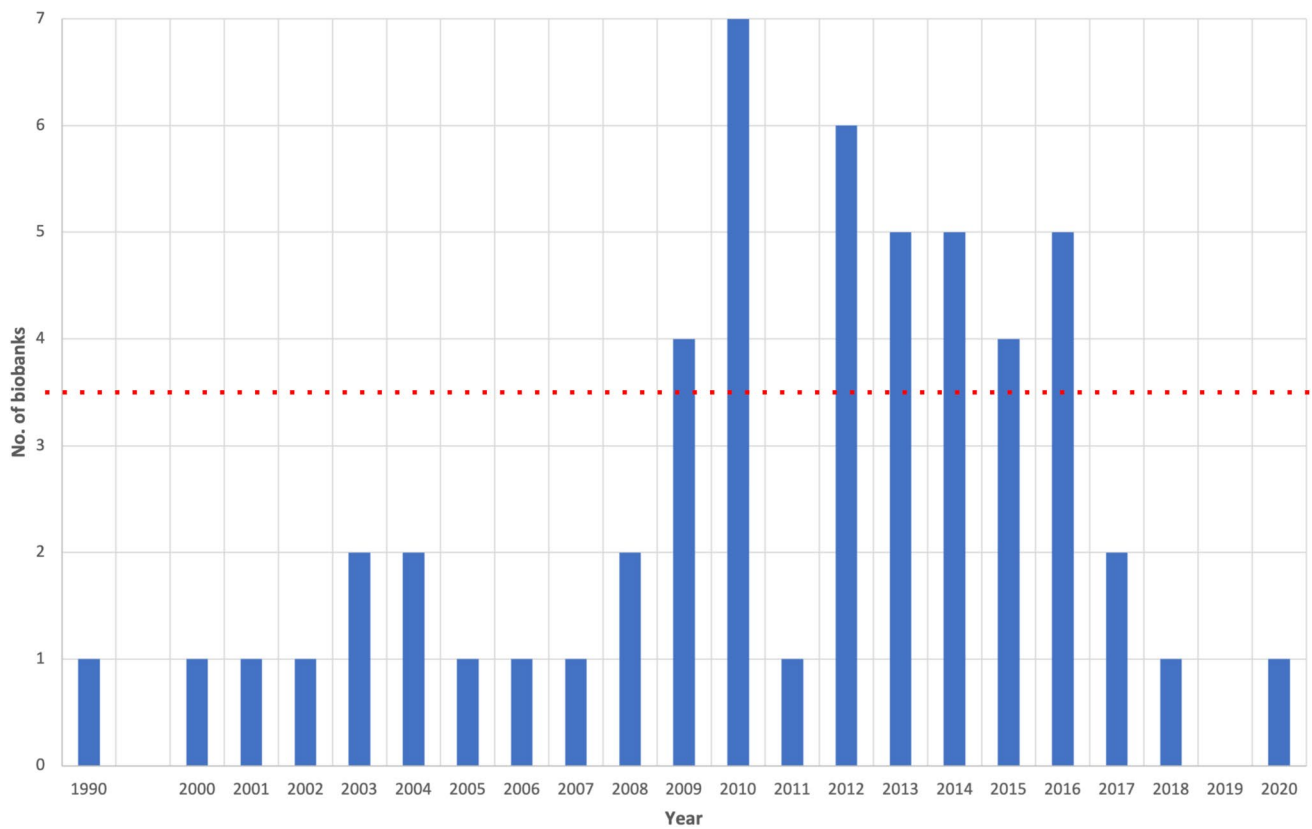
Fig. 4 Time distribution of imaging biobanks by year of development (defined as year of biobank creation, or alternatively, of beginning of patient recruitment, based on available information). The red dashed line marks 50% of the maximum number of biobanks developed in a single year ($n = 7$ in the year 2010)

extent to which type 2 diabetes contributes to its development and progression. In this way, the project is expected to deliver biological markers that could indicate which diabetic patients are at a greater risk for developing diabetic cardiomyopathy, possibly leading to a more detailed understanding of the disease [78]

ImaLife is a population-based project aimed to investigate early imaging biomarkers of three common diseases in the general population (i.e. cardiovascular disease, lung cancer and COPD), in an effort to prompt earlier treatment and reduce mortality [30].

The integration of quantitative features extracted from medical images and associated clinical data is a key factor to catalyse a change of paradigm in precision medicine healthcare, to assist diagnosis and prognosis, to predict valid disease outcomes and to understand the mechanisms behind complex diseases by learning from retrospective data [79]. Nevertheless, most results emerging from radiomic studies are negatively affected by a reduced number of cases collected, and novel biomarkers need to be validated in large multicentre populations. A small sample size can be a critical shortcoming for finding statistically significant correlations between treatment outcomes and radiomics features

with high confidence intervals [80, 81]. This problem is especially significant when the different data repositories that form biobanks are intended to be used as a basis for the generation of predictive models (i.e. diagnosis, prognosis or response to treatment). In this case, the sample size of the datasets is crucial for the models developed to be valid and robust. Sample size estimation is essential to ensure that the results are conclusive and representative of the cohort under investigation, as well as to increase the degree of reproducibility of the study generated from the repository data.

In light of the above, the design of a biobank oriented towards exploitation using radiomics and AI predictive models should consider that the constructed data repository must have sufficient cases corresponding to the different events that may have been experienced by the subjects enrolled in a biobank [82]. Interestingly, while in the previous decade only pathological and clinical data were exploited for making predictions of disease outcomes, nowadays genomic, proteomic and imaging data extracted by high-throughput screening techniques are available for most patients with several diseases, including cancer [83].

Oncology is a suitable field for the discovery of imaging biomarkers, since cancer patients are frequently

monitored with different imaging modalities for staging and post-treatment follow-up. In this setting, the correlation of imaging data with genetic, pathological, clinical and molecular data is of paramount importance to allow the development of predictive models [80].

The quantity and representativeness of the datasets used and the availability of high-performance computing infrastructures (HPCI) can have a strong impact on the validity of the proposed model of imaging biobank [84, 85]. Moreover, the standardisation process ensures data quality and the interoperability of different data sources through shared message formats and controlled terminologies, thus fostering the collaboration between biobanks and interoperability in data sharing [84]. As an indispensable requirement for research projects, the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles must be met, ensuring that data are traceable, accessible, interoperable and reusable, which has a direct impact on the definition of the biobank architecture. Standards like the Minimum Information About BIobank data Sharing (MIABIS) [86] or the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) [87] can facilitate the exchange of sample information and data and the systematic analysis of heterogeneous observational databases, making the biobank compliant with the specified FAIR principles. To this regard, an integration model of imaging data and biological data expanding the MIABIS standard with DICOM metadata has recently been proposed [88].

The present study has one limitation in that not all imaging biobanks could be traced back to specific literature articles, partly because several biobanks are still under construction. Another limitation is the difficulty of finding information regarding the biobank (for instance, many biobanks do not have a website) and the way imaging data were handled.

In conclusion, existing imaging biobanks can serve as infrastructures aimed at collecting radiological images, which could be used to obtain radiomic data and to train and validate AI models. Links between biobanks (either pathological, imaging or clinical) are needed to fuel AI and radiomics research, aiding the development of precision medicine and promoting earlier transfer into clinical practice.

Declarations

Guarantor The scientific guarantor of this publication is Prof. Emanuele Neri.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not necessary because our manuscript is a literature review on existing imaging biobanks

Ethical approval Institutional Review Board approval was not required because our manuscript is a literature review.

Methodology
● Review paper

References

1. Litton J-E (2018) Launch of an infrastructure for health research: BBMRI-ERIC. Biopreserv Biobank 16:233–241
2. http://biospecimens.cancer.gov/patientcorner/. Accessed 12 Aug 2021
3. Siwek M (2015) An overview of biorepositories—past, present, and future. Mil Med 180:57–66
4. Marodin G, França P, da Rocha JCC, Campos AH (2012) Biobanking for health research in Brazil: present challenges and future directions. Rev Panam Salud Publica 31:523–528
5. Coppola L, Cianflone A, Grimaldi AM et al (2019) Biobanking in health care: evolution and future directions. J Transl Med 17:172
6. WMA Declaration of Taipei on ethical considerations regarding health databases and biobanks. https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks. Accessed 12 Aug 2021
7. Ikram MA, Arfan Ikram M, van der Lugt A et al (2015) The Rotterdam Scan Study: design update 2016 and main findings. Eur J Epidemiol 30:1299–1315
8. Bild DE (2002) Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 156:871–881
9. Littlejohns TJ, Holliday J, Gibson LM et al (2020) The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun 11:2624
10. European Society of Radiology (ESR) (2015) ESR position paper on imaging biobanks. Insights Imaging 6:403–410
11. Neri E, Regge D (2017) Imaging biobanks in oncology: European perspective. Future Oncol 13:433–441
12. Neri E, Del Re M, Paiar F et al (2018) Radiomics and liquid biopsy in oncology: the holons of systems medicine. Insights Imaging 9:915–924
13. https://www.primageproject.eu/. Accessed 12 Aug 2021
14. https://chaimeleon.eu/. Accessed 12 Aug 2021
15. https://www.procancer-i.eu. Accessed 12 Aug 2021
16. https://eucanimage.eu/. Accessed 12 Aug 2021
17. Macheiner T, Huppertz B, Bayer M, Sargsyan K (2017) Challenges and driving forces for business plans in biobanking. Biopreserv Biobank 15:121–125

18. Mendy M, Caboux E, Sylla BS et al (2014) Infrastructure and facilities for human biobanking in low- and middle-income countries: a situation analysis. Pathobiology 81:252–260

19. Vaught J, Lockhart NC (2012) The evolution of biobanking best practices. Clin Chim Acta 413:1569–1575

20. Ballantyne A, Eriksson S (2019) Research ethics revised: the new CIOMS guidelines and the World Medical Association Declaration of Helsinki in context. Bioethics 33:310–311

21. Chassang G, Rial-Sebbag E (2018) Research biobanks and health databases: the WMA Declaration of Taipei, added value to european legislation (soft and hard law). Eur J Health Law 25:501–516

22. Shabani M, Chassang G, Marelli L (2021) The impact of the GDPR on the governance of biobank research. In: Slokenberga S, Tzortzatou O, Reichel J (eds) GDPR and biobanking. Law, governance and technology series, vol 43. Springer, Cham. https://doi.org/10.1007/978-3-030-49388-2_4

23. Gatidis S, Heber SD, Storz C, Bamberg F (2017) Population-based imaging biobanks as source of big data. Radiol Med 122:430–436

24. Moher D (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151:264

25. Malden D, Lacey B, Emberson J et al (2019) Body fat distribution and systolic blood pressure in 10,000 adults with whole-body imaging: UK Biobank and Oxford BioBank. Obesity (Silver Spring) 27:1200–1206

26. Al Kuwari H, Al Thani A, Al Marri A et al (2015) The Qatar Biobank: background and methods. BMC Public Health 15:1208

27. Casasnovas JA, Alcaide V, Civeira F et al (2012) Aragon workers' health study – design and cohort description. BMC Cardiovasc Disord 12:45

28. Ikram MA, Brusselle GGO, Murad SD et al (2017) The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 32:807–850

29. Vonder M, van der Aalst CM, Vliegenthart R et al (2018) Coronary artery calcium imaging in the ROBINSCA trial: rationale, design, and technical background. Acad Radiol 25:118–128

30. Xia C, Rook M, Pelgrim GJ et al (2020) Early imaging biomarkers of lung cancer, COPD and coronary artery disease in the general population: rationale and design of the ImaLife (Imaging in Lifelines) Study. Eur J Epidemiol 35:75–86

31. Bamberg F, Kauczor H-U, Weckbach S et al (2015) Whole-body MR imaging in the German National Cohort: rationale, design, and technical background. Radiology 277:206–220

32. Lee C, Choe EK, Choi JM, et al (2018) Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. BMJ Open 8:e019327

33. Schumann G, Loth E, Banaschewski T et al (2010) The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry 15:1128–1139

34. Holmes AJ, Hollinshead MO, O'Keefe TM, et al (2015) Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. Sci Data 2:150031

35. Bergström G, Berglund G, Blomberg A et al (2015) The Swedish CArdioPulmonary BioImage Study: objectives and design. J Intern Med 278:645–659

36. Honda C, Watanabe M, Tomizawa R; Osaka Twin Research Group, Sakai N (2019) Update on Osaka University Twin Registry: an overview of multidisciplinary research resources and biobank at Osaka University Center for Twin Research. Twin Res Hum Genet 22:597–601

37. Jiang CQ, Lam TH, Lin JM et al (2010) An overview of the Guangzhou biobank cohort study-cardiovascular disease subcohort (GBCS-CVD): a platform for multidisciplinary collaboration. J Hum Hypertens 24:139–150

38. Berge T, Vigen T, Pervez MO et al (2015) Heart and brain interactions–the Akershus Cardiac Examination (ACE) 1950 Study Design. Scand Cardiovasc J 49:308–315

39. Fernández-Ortiz A, Jiménez-Borreguero LJ, Peñalvo JL et al (2013) The Progression and Early detection of Subclinical Atherosclerosis (PESA) study: rationale and design. Am Heart J 166:990–998

40. Anand SS, Abonyi S, Arbour L et al (2018) Canadian Alliance for Healthy Hearts and Minds: first nations cohort study rationale and design. Prog Community Health Partnersh 12:55–64

41. Kruithof CJ, Kooijman MN, van Duijn CM et al (2014) The Generation R Study: biobank update 2015. Eur J Epidemiol 29:911–927

42. Walker L, Chang L-C, Nayak A et al (2016) The diffusion tensor imaging (DTI) component of the NIH MRI study of normal brain development (PedsDTI). Neuroimage 124:1125–1130

43. Casey BJ, Cannonier T, Conley MI et al (2018) The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. Dev Cogn Neurosci 32:43–54

44. Harms MP, Somerville LH, Ances BM et al (2018) Extending the Human Connectome project across ages: imaging protocols for the lifespan development and aging projects. Neuroimage 183:972–984

45. Sidorenkov G, Nagel J, Meijer C et al (2019) The OncoLifeS databiobank for oncology: a comprehensive repository of clinical data, biological samples, and the patient's perspective. J Transl Med 17:374

46. Reijs BLR, Teunissen CE, Goncharenko N et al (2015) The central biobank and virtual biobank of BIOMARKAPD: a resource for studies on neurodegenerative diseases. Front Neurol 6:216

47. Kott KA, Vernon ST, Hansen T, et al (2019) Biobanking for discovery of novel cardiovascular biomarkers using imaging-quantified disease burden: protocol for the longitudinal, prospective, BioHEART-CT cohort study. BMJ Open 9:e028649

48. Schram MT, Sep SJS, van der Kallen CJ et al (2014) The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. Eur J Epidemiol 29:439–451

49. Ferreira LE, de França PHC, Nagel V et al (2017) Joinville stroke biobank: study protocol and first year's results. Arq Neuropsiquiatr 75:881–889

50. Knudsen GM, Jensen PS, Erritzoe D et al (2016) The Center for Integrated Molecular Brain Imaging (Cimbi) database. Neuroimage 124:1213–1219

51. Niklinski J, Kretowski A, Moniuszko M et al (2017) Systematic biobanking, novel imaging techniques, and advanced molecular analysis for precise tumor diagnosis and therapy: the Polish MOBIT project. Adv Med Sci 62:405–413

52. Akinyemi RO, Akinwande K, Diala S et al (2018) Biobanking in a challenging African environment: unique experience from the SIREN project. Biopreserv Biobank 16:217–232

53. Alexander LM, Escalera J, Ai L, et al (2017) An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data 4:170181

54. Di Martino A, Yan C-G, Li Q et al (2014) The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 19:659–667

55. Weiner MW, Veitch DP, Aisen PS et al (2012) The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 8:S1–S68

56. van der Flier WM, Scheltens P (2018) Amsterdam Dementia Cohort: performing research to optimize care. J Alzheimers Dis 62:1091–1111

57. Clark K, Vendt B, Smith K et al (2013) The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26:1045–1057

58. Bauermeister S, Orton C, Thompson S et al (2020) The Dementias Platform UK (DPUK) Data Portal. Eur J Epidemiol 35:601–611

59. Szendroedi J, Saxena A, Weber KS et al (2016) Cohort profile: the German Diabetes Study (GDS). Cardiovasc Diabetol 15:59

60. Martí-Bonmatí L, Alberich-Bayarri Á, Ladenstein R et al (2020) PRIMAGE project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers. Eur Radiol Exp 4:22

61. Clavreul A, Soulard G, Lemée J-M et al (2019) The French glioblastoma biobank (FGB): a national clinicobiological database. J Transl Med 17:133

62. Das S, Abou-Haidar R, Rabalais H et al (2021) The C-BIG Repository: an institution-level open science platform. Neuroinformatics. https://doi.org/10.1007/s12021-021-09516-9

63. Manniën J, Ledderhof T, Verspaget HW et al (2017) The Parelsnoer Institute: a national network of standardized clinical biobanks in the Netherlands. Open Journal of Bioresources 4:3. https://doi.org/10.5334/ojb.23

64. Aibaidula A, Lu J-F, Wu J-S et al (2015) Establishment and maintenance of a standardized glioma tissue bank: Huashan experience. Cell Tissue Banking 16:271–281

65. Duncan D, Vespa P, Pitkänen A et al (2019) Big data sharing and analysis to advance research in post-traumatic epilepsy. Neurobiol Dis 123:127–136

66. Karch A, Vogelmeier C, Welte T et al (2016) The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. Respir Med 114:27–37

67. Respondek G, Höglinger GU (2021) DescribePSP and ProPSP: German multicenter networks for standardized prospective collection of clinical data, imaging data, and biomaterials of patients with progressive supranuclear palsy. Front Neurol 12:644064

68. Besser L, Kukull W, Knopman DS et al (2018) Version 3 of the National Alzheimer's Coordinating Center's uniform data set. Alzheimer Dis Assoc Disord 32:351–358

69. Maas AIR, Menon DK, Steyerberg EW et al (2015) Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. Neurosurgery 76:67–80

70. Fowler C, Rainey-Smith SR, Bird S et al (2021) Fifteen years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) Study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. J Alzheimers Dis Rep 5:443–468

71. Elbers DC, Fillmore NR, Sung F-C, et al (2020) The Veterans Affairs Precision Oncology Data Repository, a clinical, genomic, and imaging research database. Patterns 1:100083

72. Baugh J, Bartels U, Leach J et al (2017) The international diffuse intrinsic pontine glioma registry: an infrastructure to accelerate collaborative research for an orphan disease. J Neurooncol 132:323–331

73. Ofori E, Du G, Babcock D et al (2016) Parkinson's disease biomarkers program brain imaging repository. Neuroimage 124:1120–1124

74. Rosenthal A, Gabrielian A, Engle E et al (2017) The TB Portals: an open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. J Clin Microbiol 55:3267–3282

75. Schwedt TJ, Digre K, Tepper SJ et al (2020) The American Registry for Migraine Research: research methods and baseline data for an initial patient cohort. Headache 60:337–347

76. Giese A-K, Schirmer MD, Donahue KL, et al (2017) Design and rationale for examining neuroimaging genetics in ischemic stroke: the MRI-GENIE study. Neurol Genet 3:e180

77. Lakhani DA, Chen S-C, Antic S et al (2021) Establishing a cohort and a biorepository to identify biomarkers for early detection of lung cancer: the Nashville Lung Cancer Screening Trial Cohort. Ann Am Thorac Soc 18:1227–1234

78. https://cardiateam.eu/. Accessed 12 Aug 2021

79. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=4188. Accessed 12 Aug 2021

80. Martí-Bonmatí L, Ruiz-Martínez E, Ten A, Alberich-Bayarri A (2018) Cómo integrar la información cuantitativa en el informe radiológico del paciente oncológico. Radiologia 60:43–52

81. Shi L, He Y, Yuan Z et al (2018) Radiomics for response and outcome assessment for non-small cell lung cancer. Technol Cancer Res Treat 17:1533033818782788

82. Riley RD, Ensor J, Snell KIE, et al (2020) Calculating the sample size required for developing a clinical prediction model. BMJ 368:m441

83. Wu P-Y, Cheng C-W, Kaddi CD et al (2017) Omic and electronic health record big data analytics for precision medicine. IEEE Trans Biomed Eng 64:263–273

84. Im K, Gui D, Yong WH (2019) An introduction to hardware, software, and other information technology needs of biomedical biobanks. Methods Mol Biol 1897:17–29

85. Norlin L, Fransson MN, Eriksson M et al (2012) A minimum data set for sharing biobank samples, information, and data: MIABIS. Biopreserv Biobank 10:343–348

86. Merino-Martinez R, Norlin L, van Enckevort D et al (2016) Toward global biobank integration by implementation of the minimum information about biobank data sharing (MIABIS 2.0 Core). Biopreserv Biobank 14:298–306

87. https://www.ohdsi.org/data-standardization. Accessed 12 Aug 2021

88. Scapicchio C, Gabelloni M, Forte SM et al (2021) DICOM-MIABIS integration model for biobanks: a use case of the EU PRIMAGE project. Eur Radiol Exp 5:20

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.