

METHODOLOGY

Open Access

DICOM-MIABIS integration model for biobanks: a use case of the EU PRIMAGE project



Camilla Scapicchio^{1*} , Michela Gabelloni¹, Sara Maria Forte¹, Leonor Cerdá Alberich², Lorenzo Faggioni¹, Rita Borgheresi¹, Paola Erba¹, Fabiola Païar¹, Luis Marti-Bonmati³ and Emanuele Neri¹

Abstract

PRIMAGE is a European Commission-financed project dealing with medical imaging and artificial intelligence aiming to create an imaging biobank in oncology. The project includes a task dedicated to the interoperability between imaging and standard biobanks. We aim at linking Digital imaging and Communications in Medicine (DICOM) metadata to the Minimum Information About Biobank data Sharing (MIABIS) standard of biobanking. A very first integration model based on the fusion of the two existing standards, MIABIS and DICOM, has been developed. The fundamental method was that of expanding the MIABIS core to the imaging field, adding DICOM metadata derived from CT scans of 18 paediatric patients with neuroblastoma. The model was developed with the relational database management system Structured Query Language. The integration data model has been built as an Entity Relationship Diagram, commonly used to organise data within databases. Five additional entities have been linked to the “Image Collection” subcategory in order to include the imaging metadata more specific to the particular type of data: Body Part Examined, Modality Information, Dataset Type, Image Analysis, and Registration Parameters. The model is a starting point for the expansion of MIABIS with further DICOM metadata, enabling the inclusion of imaging data in biorepositories.

Keywords: Biological specimen banks, Database management systems, Neuroblastoma, Picture archiving and communication system, Radiology

Key points

- The availability of imaging biobanks supports tumour decision-making in precision medicine.
- An integration model linking the Minimum Information About Biobank data Sharing and Digital imaging and Communications in Medicine standards has been developed.
- It enables the inclusion of imaging and biomarker data into standard biobanks.

Background

The availability of biobanks and biorepositories for the research community allows the access to multiple types of data and is the basis for big data analytics in the medical domain [1]. The term biobanking refers to the process by which massive collections of biological material and associated information are collected for research use [2]. Biobanks are defined as repositories for the storage and retrieval of samples of bodily fluids or tissues and accompanying data of the related subject, supporting contemporary research such as genomics and personalised medicine [3, 4].

Digital innovation in the healthcare systems is evolving toward a data-driven science, making high-quality data and information exchange essential in healthcare. It is

* Correspondence: camilla.scapicchio@med.uniipi.it

¹Department of Translational Research on New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
Full list of author information is available at the end of the article

therefore clear that advances are required in data management to facilitate the development of platforms for more effective usage of large volumes of data, and biobanks play a central role in this scenario [5, 6].

The European Biobanking and BioMolecular resources Research Infrastructure (BBMRI-ERIC) has been created with the aim to harmonise biobanks across Europe and sustain a network of research [7]. The European Society of Radiology (ESR) started a collaboration with BBMRI-ERIC in 2014, recognising the importance of integrating imaging and “omics” data.¹ All biobanks are indexed in the BBMRI-ERIC Directory from which researchers across Europe can seek retrieval of samples/data, or collect/host services for their samples/data [8, 9].

To integrate different biobanks, the research infrastructure uses a standard for data description named Minimum Information About Biobank data Sharing (MIABIS) 2.0 Core, as a data model standard to describe samples and sample donors. The MIABIS is a recommendation about what information should be stored in biobank to facilitate the exchange of samples information and data. MIABIS represents the standard for the interoperability between biobanks [10, 11].²

To date, biobanks are mainly focused on data describing fluid or tissue samples, and very few in BBMRI-ERIC network include image collections [12]. A novel field of research is that of imaging biobanks [13], defined as “organized databases of medical image collections associated with imaging biomarkers” [9]. Imaging biobanks store image collections, just as standard biobanks store biological samples, and donors of images can be described in the same way as donors of samples. Most imaging biobanks focus on the collection of cancer-related data and oncologic imaging biomarkers [14, 15].

A recent challenge is that of reliably connecting the available imaging biobanks to tissue biobanks to explore imaging biomarkers. In fact, the integration of molecular and imaging data is needed for a radiogenomic approach to the patient in a personalised medicine setting [16]. Linkage and integration of existing image data repositories as well as the connection between imaging biobanks and traditional biobanks is the key factor to develop and validate new imaging biomarkers, as well as to improve the general understanding of their biological significance [17].

It is therefore necessary to identify a model of interoperability that describes medical imaging data into the traditional biobanks. To reach this goal a specific task of interoperability has been designed by the PRedictive In-

silico Multiscale Analytics to support cancer personalised diagnosis and prognosis, Empowered by imaging biomarkers (PRIMAGE) project [18, 19]. PRIMAGE is a European Commission-financed project dealing with medical imaging and artificial intelligence aiming to create an imaging biobank in oncology. As part of the interoperability task, the aim of this study was to expand the MIABIS standard with DICOM metadata and provide a model of integration of imaging data and biobank samples.

Project overview

Here follows a theoretical overview of the method adopted to develop the metadata model that allows the description of imaging data and facilitate interoperability between imaging biomarkers and the traditional biobanks.

To build this model of interoperability among heterogeneous data, the existing formats and ontologies for image and data description, the MIABIS and the DICOM, were considered as standard of reference [20–22]. The MIABIS is the standard that describes the biological samples [10, 11].

In the MIABIS core, there are three main well defined components on which the other entities depend: biobank, samples collection, and study (Fig. 1) [10]. MIABIS has a modular structure that allows to add components in a flexible way, which makes the core applicable for a variety of use cases domains of biobanking and research studies. It is possible to attach additional components to the core in order to describe particular subdomains, by merely introducing auxiliary entities in the Entity-Relationship Diagram (ERD) representing the data model.

As it is well known in the imaging domain, DICOM is the standard for the visualisation and sharing of medical images and associated information [23]. In clinical practice, a medical image is a DICOM file. It includes a header in addition to the actual image [24, 25]. All the information stored in the header is cataloged in groups of elements, called “DICOM tags.” Real-world entities (e.g., images, procedures, or interpretation reports) are represented in the DICOM semantic data model by templates of attributes or data elements, and each tag identifies an attribute [23, 26].

Therefore, the integration data model here proposed is based on the attempt to create a linkage between the MIABIS format and the DICOM format to represent imaging biomarkers of the PRIMAGE biobank into a network of traditional biobanks.

Study protocol

The fundamental method used to create the model is based on a fusion between the two existing formats,

¹The last author of this paper has been chair of the Imaging Biobanks working group of the European Society of Radiology.

²The first (corresponding) author of this paper is also member of the MIABIS working group.

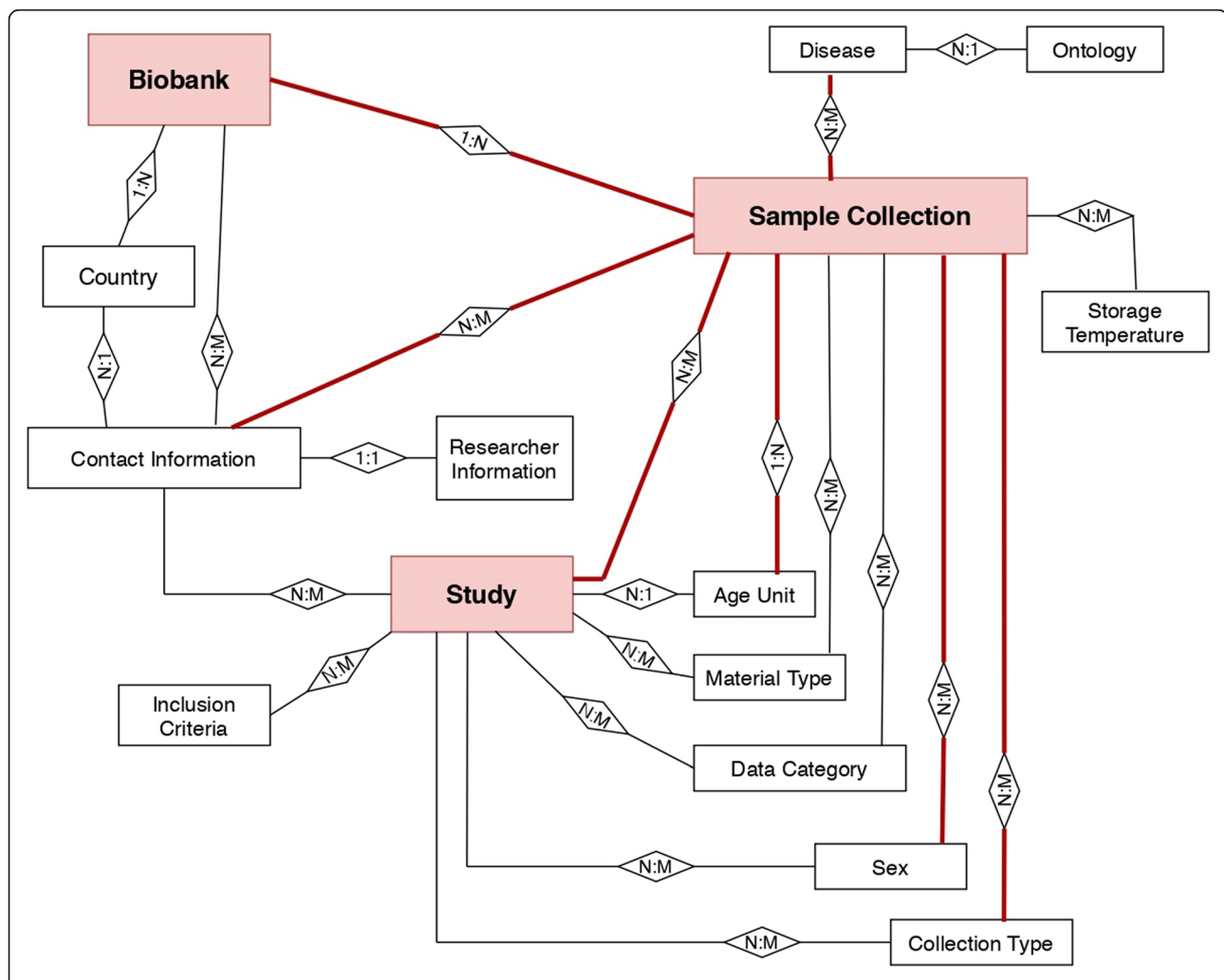


Fig. 1 Conceptual schema of the MIABIS core. The conceptual schema and the logical foundation of the MIABIS core used to summarise the data organisation. In the red boxes, the three main categories biobank, study, and sample collection are represented, with the relationships between them and the sub-entities linked to them. The relationships highlighted in red are the ones that have been maintained in the new integration model but linked to "Collection" in place of "Sample Collection." The abbreviations "1:1," "1:N," and "N:M" define the type of relationship in the relational database design. The one-to-one (1:1) relationship exists when zero or one instance of entity A can be associated with zero or one instance of entity B, and vice-versa. The one-to-many (1:N) relationship exists when, for one instance of entity A, there exists zero, one, or many instances of entity B; but for one instance of entity B, there exists zero or one instance of entity A. A many-to-many (M:N) relationship exists when, for one instance of entity A, there exists zero, one, or many instances of entity B, and vice-versa

DICOM and MIABIS standards, mainly describing imaging collections and sample tissue data collections respectively.

The integration data model has been built as an Entity Relationship Diagram (ERD), commonly used to organise data within databases. It is the logical data scheme, which describes the logical organisation and structure of data contained in the database, in a formal language supported by a database management system (DBMS). In a relational database, the scheme defines the given attributes of each table/entity and the relationships between tables. The steps followed to create this model and build the diagram are explained in more detail in the following.

The strategy has been to maintain the fundamental core of the MIABIS data model as a starting point. The conceptual schema and the logical foundation of the ERD describing the MIABIS core are represented in Fig. 1, where the three main categories: biobank, study, and sample collection and the relationships between them and sub-entities have been drawn. In terms of database implementation, the new integration model has been built starting from the MIABIS Entity Relationship Diagram (ERD) coming from the conceptual schema. The tables representing the three main entities with their attributes and the sub-entities linked to them through specific relationships, as represented in the MIABIS conceptual schema (Fig. 1), have been used as a starting

point and left almost unchanged in the new extended model, without too many alterations. The MIABIS 2.0 core is modularised and flexible; this means that it is possible to attach additional components to it. We took advantage of this property, *i.e.*, we precisely added new components to the core to obtain the new model applicable to the use case of imaging data.

The step necessary to perform the extension of this starting model to the imaging field has been the replacement of the “Sample Collection” entity with a more generic macro category called “Collection,” which includes multiple types of data represented by several sub-categories added to this macro category, *i.e.*, “Image Collection,” “Sample Tissue Collection,” “Clinical Variable Collection,” “Molecular Biomarker Collection,” and “Software/Source Code” entities. This is the fundamental step that enables a merge among different types of collections.

The extension of the model to the imaging data has been then done by expanding the “Image Collection” sub-category in two main parts. On the one hand, there is the part regarding the imaging metadata more specific to the particular type of acquired data and based on the DICOM tags. On the other hand, there is a part dependent on the process of image analysis conducted to extract imaging biomarkers and radiomic data. In fact, since the core content of imaging biobanks not only exists out of images but also of any other data (biomarkers) that may be extracted from images through computational analysis, in addition to the classes which represent the information related to imaging data and based on DICOM tags, another part of the model has been developed with other classes describing imaging biomarkers and radiomic data.

At the time of the development of this first model, the imaging data had not yet been uploaded to the PRIMAGE platform. Thus, the model development was not based on how a PRIMAGE imaging biobank sample looks like. Instead, we used the metadata of eighteen DICOM files from cases collected by our Hospital network as examples. In particular, we visualised the DICOM tags of some anonymised multislice CT studies, obtained in paediatric patients with neuroblastoma, available in the PACS system of the Pisa University Hospital, just to perform a preliminary study on which DICOM metadata are worth using for the description of these kind of imaging data. The DICOM tags have been easily visualised using the *Horos* DICOM viewer integrated in the hospital PACS system. The methodology followed to build up this model included also the close collaboration with the other partners of the project and the different professional figures involved in it. Within the PRIMAGE project, researchers at La Fe University and Polytechnic Hospital and Health Research Institute

in Valencia developed the radiomics process to extract the imaging biomarkers from neuroblastoma imaging data. Consequently, for the classes of the model describing imaging biomarkers and radiomic data, we referred to the image analysis process performed by this partner. As the project is ongoing, the analysis is described in an internal document, so far available only to the project partners. In this description of the radiomics process, we identified the main elements that can be translated as attributes to properly represent the data coming from the radiomics analysis in the model.

Software tools

The diagram has been created using the Structured Query Language (MySQL) software [27]. In particular, the MySQL Workbench, which is a graphical tool for working with MySQL servers and databases, has been used to design the database model [28]. This tool enables the graphical creation and manipulation of a model establishing relationships between the tables, where each table represents a single entity. Once defined the ERD, through this tool, it will also be possible to generate the corresponding SQL script.

Model description

The conceptual schema of the expansion is represented in Fig. 2, where the red relationships are exactly the same as the ones in the “Sample Collection” entity in MIABIS conceptual schema in Fig. 1, with the “Collection” entity being in place of the “Sample Collection” entity. The “Sample Tissue Collection” in this new schema corresponds to the same entity of the replaced “Sample Collection” in Fig. 1.

The complete ERD of the resulting model with the imaging part linked to the MIABIS core is represented in Fig. 3. This is the structural diagram manually drawn in MySQL Workbench and typically used in informatics to visualise the database design idea, with the tables representing the entities, with a set of attributes describing each entity, and the links representing the relationships among the entities.

This is the model that allows defining a unified view and description of data and will provide access to data themselves (also imaging data, in this case) and to the related clinical information through a simple query processing by researcher users, even though data are heterogeneous. We also split the entire model in Fig. 3 into Figs. 4 and 5 with a zoom on the most relevant parts of the model, the one describing the DICOM metadata and the one describing the process of image analysis for image biomarkers extraction.

In this new extended final model (Fig. 3), based on the schematically simplified expansion represented in Fig. 2, it is evident that among the sub-entities that were in

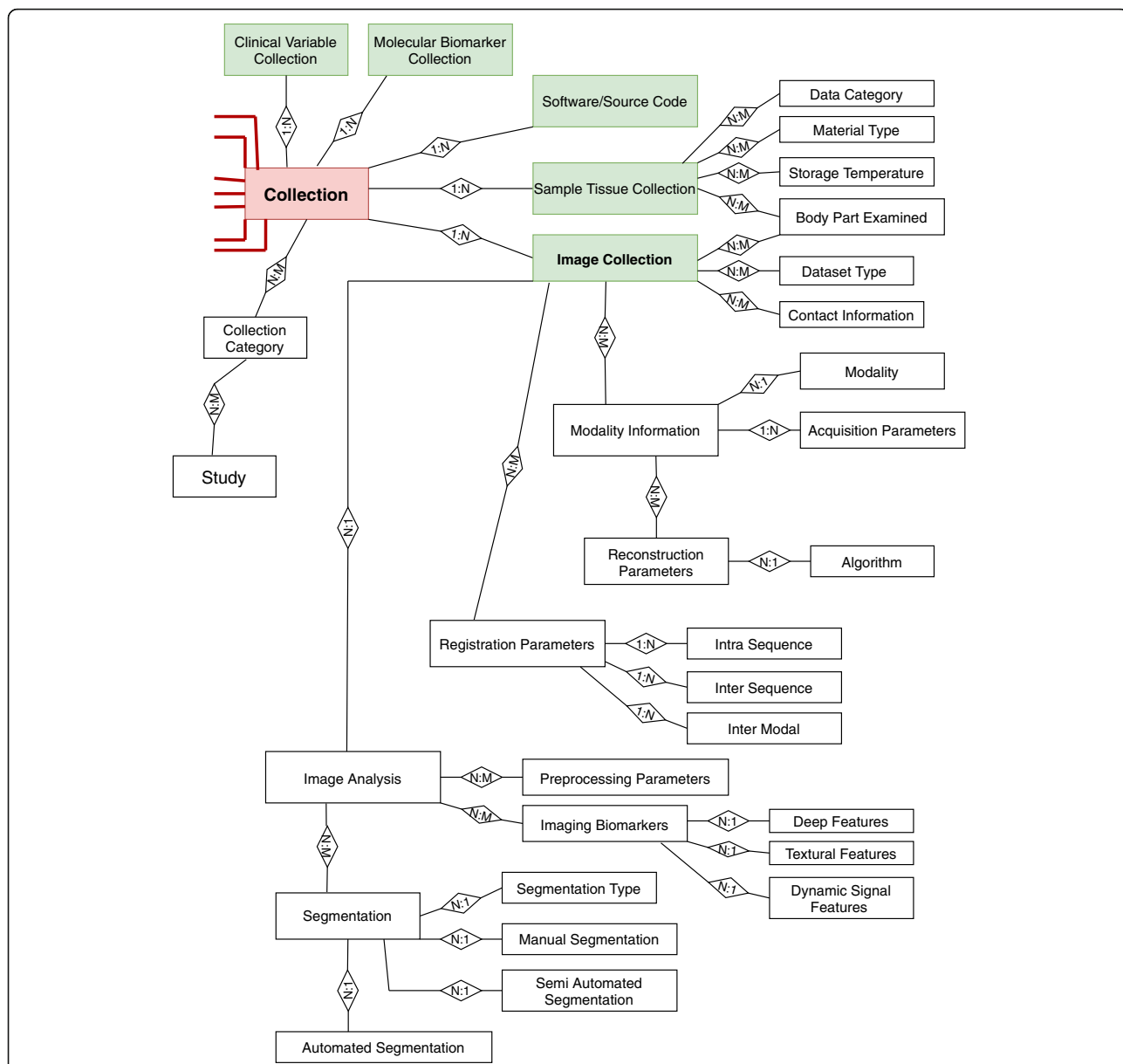


Fig. 2 Conceptual schema of the expansion of the model to image collection. The conceptual schema and the logical foundation of the expansion to image collections. The “Collection” entity replaces the “Sample Collection” entity in the conceptual schema of the MIABIS core. The red links are the same ones that connect the “Sample Collection” table to the other tables represented in Fig. 1, with the only difference that the more generic “Collection” replaces “Sample Collection.” The boxes in green are the ones representing the new sub-entities

relationship with the “Sample Collection” in the MIABIS ERD, those specifically related to tissue samples, such as storage temperature and material type, have only been linked to the “Sample Tissue Collection” sub-entity. However, the more generic sub-entities, *i.e.*, those describing patient’s information, such as age and sex, have been obviously linked to the more generic macro category “Collection.”

Five additional entities are linked to the “Image Collection” sub-category in order to include the

imaging metadata more specific to the particular type of data: body part examined, modality information, dataset type, image analysis and registration parameters [23, 29, 30]. This expansion is the fundamental step that enables to extend the MIABIS model, only describing tissue sample data, to the imaging data.

The Body Part Examined entity is based on its respective DICOM tag (0018, 0015) related to DICOM Part 16 which provides a list of 116 terms describing with a text

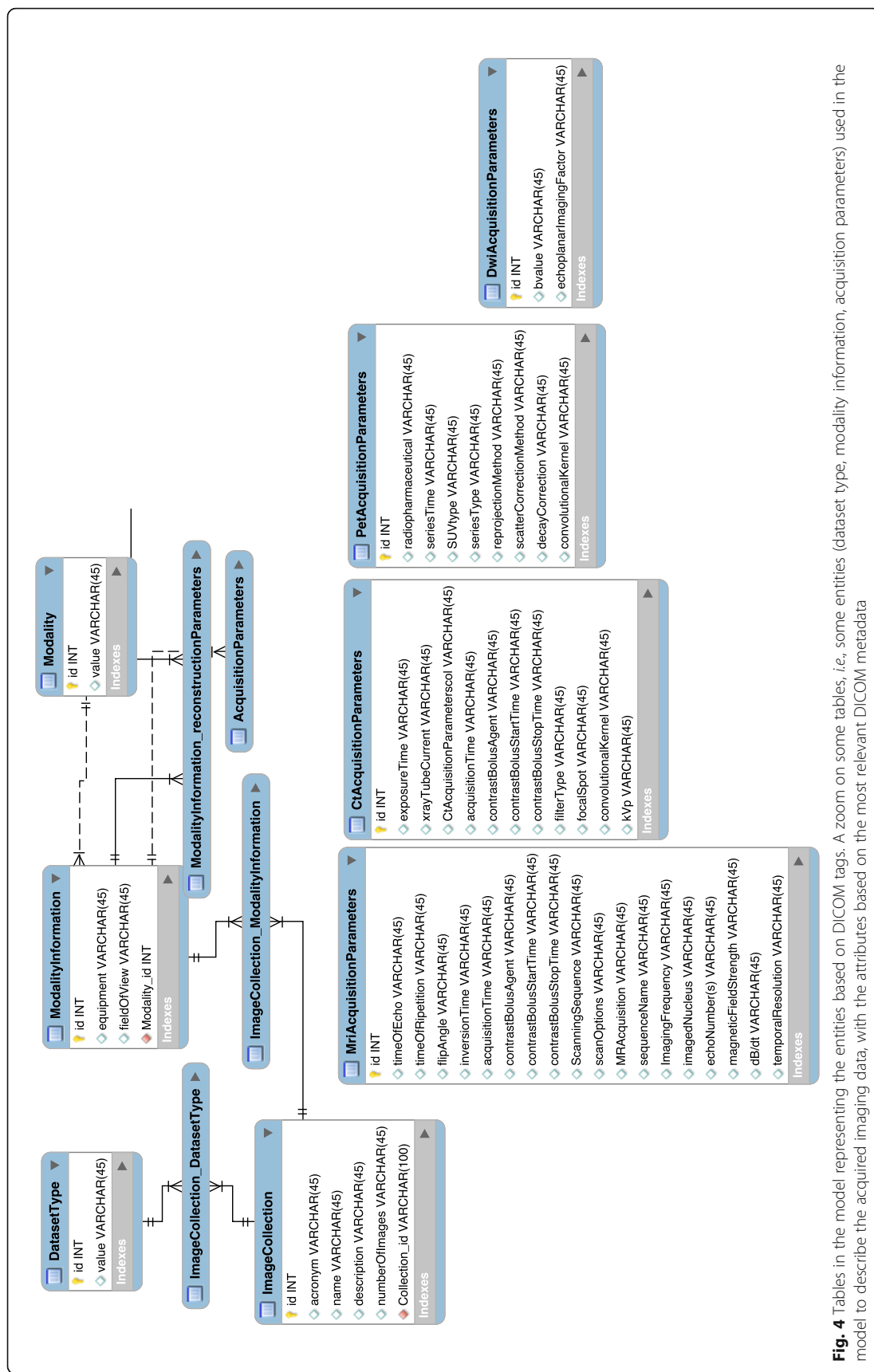


Fig. 4 Tables in the model representing the entities based on DICOM tags. A zoom on some tables, i.e., some entities (dataset type, modality information, acquisition parameters) used in the model to describe the acquired imaging data, with the attributes based on the most relevant DICOM metadata

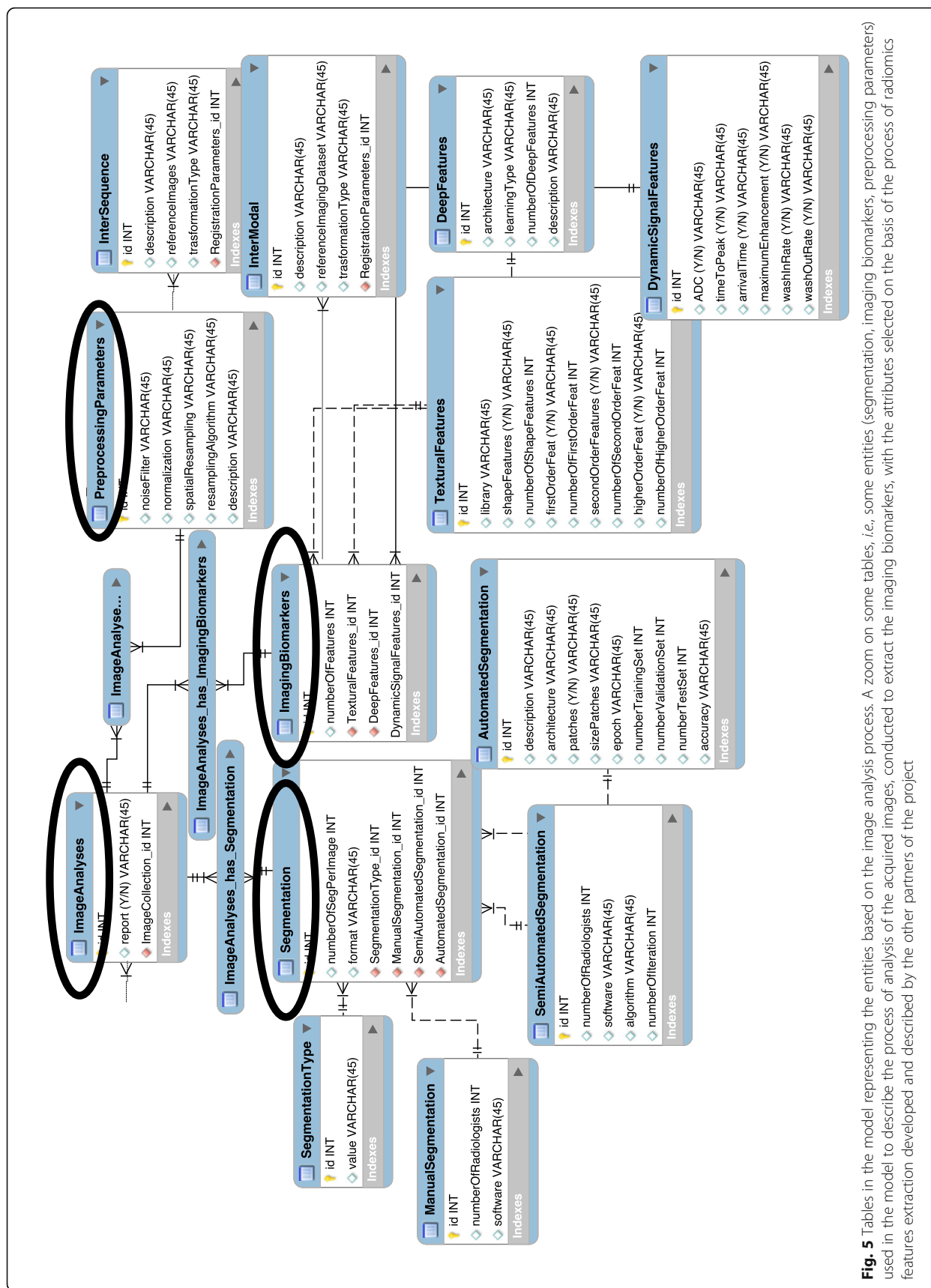


Fig. 5 Tables in the model representing the entities based on the image analysis process. A zoom on some tables, i.e., some entities (segmentation, imaging biomarkers, preprocessing parameters) used in the model to describe the process of analysis of the acquired images, conducted to extract the imaging biomarkers, with the attributes selected on the basis of the process of radiomics features extraction developed and described by the other partners of the project

Clinical variables

← Back to studies

PRIMAGE
Medical imaging
Artificial intelligence
Childhood cancer research

Patient data	Diagnosis	Treatment and evaluation	Follow-up
Associated symptoms	Laboratory	Imaging	Staging
Diagnosis date (Biopsy date)	<input type="text"/>	Age at diagnosis (months)	<input type="text"/>
First symptoms date	<input type="text"/>		
LTS (Life Threatening Symptoms)			
Intraspinal neuroblastoma	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Pain requiring opiate treatment	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Vomiting needing nasogastric/IV support	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Weight loss > 10% body weight	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Respiratory distress without evidence of infection	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Hypertension	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
IVC (inferior vena cava) compression ± leg edema	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Creatinine > 2x Upper Limit Normal	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Urine output < 2 mL/Kg/day	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Hydrourter / hydronephrosis	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Abnormal liver function > 2x Upper Limit Normal	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Evidence of DIC (Disseminated Intravascular Coagulation)	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Platelets < 50 x 10 ⁹ /L	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Bladder/bowel dysfunction secondary to mass effect	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
Very large tumor volume causing concern of possible tumor rupture and/or the possible rapid development of systemic upset	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown		

Fig. 6 Example of an e-form representing some of the clinical variables related to a PRIMAGE sample. Detailed information about the associated symptoms at the time of diagnosis in the diagnosis section of clinical variables in an e-form. This is an example of how it appears in the PRIMAGE platform

the part of the body examined. It may also be replaced by Anatomic Region Sequence DICOM tag (0008, 2218).

The dataset type entity is likewise based on its corresponding DICOM tag (0008, 0016) related to DICOM Part 16 that provides a list of 120 terms and code values. The Modality Information entity describes the

equipment that originally acquired the data used to create the images in the series and is in relationship with modality, acquisition parameters, and reconstruction parameters. Modality is based on the DICOM tag (0008, 0060) related to DICOM Part 03 which provides a list of 54 terms and code values.

Clinical variables

← Back to studies

PRIMAGE
Medical imaging
Artificial intelligence
Childhood cancer research

Patient data	Diagnosis	Treatment and evaluation	Follow-up
Associated symptoms	Laboratory	Imaging	Staging
LDH	<input type="checkbox"/> Evaluated <input type="checkbox"/> Not evaluated		
Pathology tumor	<input type="checkbox"/> Evaluated <input type="checkbox"/> Not evaluated		
Pathology bone marrow	<input type="checkbox"/> Evaluated <input type="checkbox"/> Not evaluated		
Liquid biopsy (peripheral blood)	<input type="checkbox"/> Evaluated <input type="checkbox"/> Not evaluated		
Molecular studies	<input type="checkbox"/> Evaluated <input type="checkbox"/> Not evaluated		

Save

Fig. 7 Example of part of an e-form representing some clinical variables related to a PRIMAGE sample. Detailed information about the laboratory tests in the diagnosis section of clinical variables in an e-form. This is an example of how it appears in the PRIMAGE platform

With regard to this part of the schema based on the DICOM metadata, only some attributes describing the various entities have been established. In Fig. 4, we show some tables from the entire model in Fig. 3, with the attributes based on the selected DICOM tags. In the future development of the model, a more accurate examination and analysis on the meaningful DICOM tags that must be used as attributes to describe the specific type of imaging data will be further explored.

In the part concerning the modality information, the image analysis, and the registration parameters, the tables representing the entities and the relationships between entities and sub-entities have been established on the basis of the image analysis process and the consequent biomarkers extraction implemented in the PRIMAGE project. In addition, given the description of this image analysis process, the attributes describing these entities have been selected (Fig. 5).

The registration parameters entity is linked to intra-sequence, inter-sequence, and inter-modal sub-entities. The image analysis entity has pre-processing parameters, segmentation, and imaging biomarkers as sub-entities. Segmentation is then divided into three other sub-entities which describe the segmentation modality: manual, semi-automatic, and automatic. Imaging biomarkers is instead divided into the type of extracted features: texture feature, deep features (both related to the radiomics analysis), and dynamic signal features.

The use-case of integration between these two standards, DICOM and MIABIS, is a very first model and it is a starting point for standardisation of imaging data and metadata representation for data sharing.

Discussion

It is important to remark that the purpose of this work is to provide a theoretical description of the model and, in particular, of the idea behind the construction of the integration. The described software tool was just used as a visual support to the realisation of the model, which still needs to be validated on real data. As soon as the data are available, we will have the model ready to be used to obtain practical results confirming the initial idea here described. The step of data collection is still ongoing, so we did not link the model to concrete data yet, but we built the model on the basis of the information generally known about the data to describe them.

The significant novelty of the DICOM-MIABIS integration model developed in this study is the preparation for interoperability among imaging biobanks and traditional biobanks of the BBMRI network [18].

The proposed model will bring major advancements in facilitating imaging data cross-linking with other traditional wider biorepositories. This will have significant

potential to link imaging biomarkers with other biomarkers and clinical data. The inclusion of new imaging biobanks in standard biobanks is crucial to achieve data integration, to access large amounts of high-quality data and, consequently, to foster a personalised medicine setting [31].

A key feature that a biobank must have to be included in BBMRI is to provide access to data and clinical information through a simple query processing. In fact, data from multiple sources can be heterogeneous, *i.e.*, with different structure, models, query languages, and semantics. Therefore, data integration, which consists in providing a unified view and access of multiple datasets to users, is fundamental. The described data model, based on the proposed EER diagram, enables to organise the imaging data in PRIMAGE biobank in such a way as to meet this request, integrating imaging data and related information through a unified view [32, 33].

Another aspect that makes relevant the description of this model is that in BBMRI-ERIC Directory only one imaging biobank with CT images of COVID-19 pertaining to the Netherlands node has already been included, and there does not exist a description of the data model used for the data integration [34]. As a result, imaging biobanks and their implementation into the already existing biobanks is a very recent field of research, in the literature, there is no detailed description of a database model that allows imaging data to be integrated into a network of traditional biobanks, which are based on a standard that is only limited to describing tissue samples. The novelty of the present work consists in providing a descriptive model that should offer a possible solution on how to effectively connect imaging and tissue biobanks, giving guidance on how to organise the database to include and add medical images and associated metadata and biomarkers to other types of data.

The model has been constructed referring to the use case of neuroblastoma (NB) and diffuse intrinsic pontine glioma (DIPG) tumours, because imaging and radiomic data in PRIMAGE biobank focuses on these two paediatric cancers and no neuroblastoma or DIPG biobanks already exist. However, the model could also be extended to be adapted to other solid tumours use-cases.

Within the specific framework of the PRIMAGE project, this model will contribute to provide a strategy for embedding PRIMAGE imaging biobanks into wider biobanks networks (*e.g.*, BBMRI-ERIC), but, generally, since the advancement of in-silico medicine research is intrinsically linked to the availability of curated databases, a generalisation of such an integration model could play a part in boosting precision medicine in oncology, contributing to easier access to data and their use for AI analysis, which will hopefully provide improved health outcomes in general. In fact, for a personalised

assessment of the disease imaging biomarkers alone are not sufficient, and they must be considered in conjunction with other biological data [35]. Consequently, the possibility to access biobanks including different kinds of data, *e.g.*, imaging and biological, related to the same patient could be crucial to link images and biomarker data to other omics, such as genomic profiling, metabolomics, proteomics, lab values, and clinical information. In order to create this integration of data, it is extremely important to have a database model as a precisely described tool to efficiently organise the data [17, 36].

In conclusion, the model built so far and here described is only a first proposal of integration between MIABIS and DICOM formats. This model will be further evolved by extending the descriptions of image collections in order to figure out the necessary information to describe imaging data and define the best way to represent them into biobanks. In fact, as mentioned, this first model has been built relying on the description of the image and radiomics analysis conducted by the institution responsible for the generation of imaging and radiomic data, without having access to the real data. After the development of this proposal model, the first real imaging studies and the related clinical variables have been uploaded to the e-form in two separate projects (NB and DIPG) and are now accessible through the web platform. At the moment a PRIMAGE imaging biobank sample is described in the platform through an e-form containing all the relevant clinical variables (Figs. 6 and 7), while the imaging metadata are simply contained in the imaging metadata part of the separated DICOM file.

Therefore, in a subsequent study, it should be possible to fine-tune this first model by studying which DICOM metadata should be effectively included and which particular imaging biomarkers and radiomic features are worth representing and linking to other kinds of data (*e.g.*, clinical), now relying on real uploaded data. After having explored the PRIMAGE platform data storage and having understood how the data and the related clinical variables are organised, it will be possible to further improve the integration model by adding only the attributes that will be considered necessary and the clinical biomarkers assessed as the meaningful ones.

Furthermore, MIABIS Core is currently being updated to version 3.0. The update will include additional extensions which have been identified by BBMRI-ERIC working group. Consequently, a possible future outlook could be that of modifying the integration model to adapt it to this currently being updated version 3.0 of MIABIS core.

The model here theorised is a first attempt to link the MIABIS standard with the DICOM standard, whose final aim is to simplify access to knowledge, improve interoperability, standardisation and data management, and to ensure a harmonised approach to quality assurance of data,

enabling an integration of imaging and “omics” data with external resources (other biobanks) and the creation of a structured repository for imaging data in PRIMAGE biobank to describe them to the scientific community.

The innovative aspect and the scientific relevance of this work lie in the possibility of advancement toward reuse of datasets, which is essential to accelerate the uptake of the use of big data in medicine research.

Abbreviations

BBMRI-ERIC: European Biobanking and BioMolecular resources Research Infrastructure; DICOM: Digital Imaging and COmmunications in Medicine; DIPG: Diffuse Intrinsic Pontine Glioma; ERD: Entity Relationship Diagram; MIABIS: Minimum Information About Biobank data Sharing; NB: NeuroBlastoma; PRIMAGE: PRedictive In-silico Multiscale Analytics to support cancer personalised diaGnosis and prognosis, Empowered by imaging biomarkers; SQL: Structured Query Language

Acknowledgements

PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalised diaGnosis and prognosis, empowered by imaging biomarkers) Business Place is a Horizon 2020 | RIA (Topic SC1-DTH 07-2018) project with grant agreement no: 826494.

Authors' contributions

EN initially proposed to write the manuscript. EN and CS defined the scheme of the content. CS edited the first version of the manuscript. SMF developed the first version of the model described in the manuscript. Each author contributed to the intellectual content of the manuscript. In addition, EN, MG, RB, LCA, and LM-B added comments to improve the manuscript. All authors read and approved the final version of the manuscript.

Funding

Horizon 2020 project (RIA, topic SC1-DTH-07-2018).

Availability of data and materials

Project website (<https://www.primageproject.eu/>)

Declarations

Ethics approval and consent to participate

The 18 cases retrieved from Pisa Hospital PACS system to study the metadata have been used in agreement with the Ethics Committee. The local Ethical Committee (“Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana” in Florence) approved the study for the PRIMAGE project in date 30 September 2019. The data had been anonymised before reading the metadata and the images remained inside the hospital network for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Translational Research on New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy. ²Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain. ³Medical Imaging Department, La Fe University and Polytechnic Hospital & Biomedical Imaging Research Group (GIBI230), La Fe University and Polytechnic Hospital and Health Research Institute, Valencia, Spain.

Received: 21 September 2020 Accepted: 26 March 2021

Published online: 12 May 2021

References

1. Kinkorová J, Topolčan O (2020) Biobanks in the era of big data: objectives, challenges, perspectives, and innovations for predictive, preventive, and

- personalised medicine. *EPMA J* 11:333–341. <https://doi.org/10.1007/s13167-020-00213-2>
2. Fransson MN, Rial-Sebbag E, Brochhausen M, Litton J-E (2015) Toward a common language for biobanking. *Eur J Hum Genet* 23:22–28. <https://doi.org/10.1038/ejhg.2014.45>
 3. Asslaber M, Zatloukal K (2007) Biobanks: transnational, European and global networks. *Brief Funct Genomic Proteomic* 6:193–201. <https://doi.org/10.1093/bfpg/elm023>
 4. Coppola L, Cianflone A, Grimaldi AM et al (2019) Biobanking in health care: evolution and future directions. *J Transl Med* 17 1: 172. <https://doi.org/10.1186/s12967-019-1922-3>
 5. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2:3. <https://doi.org/10.1186/2047-2501-2-3>, 1
 6. Kumar Y, Sood K, Kaul S, Vasuja R (2020) Big data analytics and its benefits in healthcare. *Big Data Anal Healthcare* 3–21. https://doi.org/10.1007/978-3-030-31672-3_1
 7. BBMRI-ERIC Home. <https://www.bbMRI-eric.eu/>. Accessed 24 Aug 2020
 8. Holub P, Swertz M, Reihls R, van Enckevort D, Müller H, Litton JE (2016) BBMRI-ERIC Directory: 515 biobanks with over 60 million biological samples. *Biopreserv Biobanking* 14:559–562. <https://doi.org/10.1089/bio.2016.0088>
 9. BBMRI-ERIC Directory. <https://directory.bbMRI-eric.eu/menu/main/app-molgenis-app-biobank-explorer/biobankexplorer>. Accessed 24 Aug 2020
 10. 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 Biobanking* 14:298–306. <https://doi.org/10.1089/bio.2015.0070>
 11. MIABIS/miabis. <https://github.com/MIABIS/miabis>. Accessed 24 Aug 2020
 12. Kulkarni S (2016) World's largest biomedical imaging study launched by UK Biobank. *Editage Insights*. <https://doi.org/10.34193/EI-A-5902>
 13. European Society of Radiology (ESR) (2015) ESR position paper on imaging biobanks. *Insights Imaging* 6:403–410. <https://doi.org/10.1007/s13244-015-0409-x>
 14. Martí-Bonmatí L, Ruiz-Martínez E, Ten A, Alberich-Bayarri A (2018) How to integrate quantitative information into imaging reports for oncologic patients. *Radiologia* 60 Suppl 1:43–52. <https://doi.org/10.1016/j.rx.2018.02.005>
 15. Neri E, Regge D (2017) Imaging biobanks in oncology: European perspective. *Future Oncol* 13:433–441. <https://doi.org/10.2217/fon-2016-0239>
 16. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
 17. Alberich-Bayarri A, Hernández-Navarro R, Ruiz-Martínez E, García-Castro F, García-Juan D, Martí-Bonmatí L (2017) Development of imaging biomarkers and generation of big data. *Radiol Med* 122:444–448. <https://doi.org/https://doi.org/10.1007/s11547-017-0742-x>, 6
 18. Martí-Bonmatí L, Alberich-Bayarri A, 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. <https://doi.org/https://doi.org/10.1186/s41747-020-00150-9>, 1
 19. Primage Project. <https://www.primageproject.eu/>. Accessed 24 Aug 2020
 20. Rubin DL, Napel S (2010) Imaging informatics: toward capturing and processing semantic information in radiology images. *Yearb Med Inform* 19(01):34–42
 21. Wang KC (2018) Standard lexicons, coding systems and ontologies for interoperability and semantic computation in imaging. *J Digital Imaging* 31: 353–360. <https://doi.org/10.1007/s10278-018-0069-8>
 22. Stanescu L, Burdescu DD, Brezovan M, Mihai CG (2011) Creating new medical ontologies for image annotation: a case study. Springer Science & Business Media, New York, Dordrecht, Heidelberg, London
 23. Bidgood WD Jr, Horii SC, Prior FW, Van Syckle DE (1997) Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc* 4:199–212. <https://doi.org/https://doi.org/10.1136/jamia.1997.0040199>, 3
 24. Mildnerberger P, Eichelberg M, Martin E (2002) Introduction to the DICOM standard. *Eur Radiol* 12:920–927. <https://doi.org/10.1007/s003300101100>
 25. Bidgood WD, Horii SC (1992) Introduction to the ACR-NEMA DICOM standard. *Radiographics* 12:345–355. <https://doi.org/10.1148/radiographics.12.2.1561424>
 26. DICOM Controlled Terminology (2020). <http://bioportal.bioontology.org/ontologies/DCM?p=summary>. Accessed 25 Aug 2020
 27. MySQL. <https://www.mysql.com/>. Accessed 25 Aug 2020
 28. Inan DI, Juita R (2011) Analysis and design complex and large data base using MySQL workbench. *Int J Comput Sci Inform Technol* 3:173–183. <https://doi.org/10.5121/ijcsit.2011.3515>
 29. Prior FW (1993) Specifying DICOM compliance for modality interfaces. *Radiographics* 13:1381–1388. <https://doi.org/10.1148/radiographics.13.6.8290731>
 30. DICOM PS3.2 2020d - Conformance. <http://dicom.nema.org/medical/dicom/current/output/pdf/part02.pdf>. Accessed 27 Aug 2020
 31. Mantarro A, Scalise P, Neri E (2017) Imaging biobanks, big data, and population-based imaging biomarkers. In: *Imaging Biomarkers: Development and Clinical Integration*. Springer, Cham, pp 153–157. https://doi.org/10.1007/978-3-319-43504-6_13
 32. Pankowski T (2016) Faceted queries in ontology-based data integration. In: *Proceedings of the 18th International Conference on Enterprise Information Systems*, pp 150–157
 33. Gagnon M (2007) Ontology-based integration of data sources. In: *2007 10th International Conference on Information Fusion*. IEEE, pp 1–8. <https://doi.org/10.1109/ICIF.2007.4408086>
 34. Holub P, Kozera L, Florindi F et al (2020) BBMRI-ERIC's contributions to research and knowledge exchange on COVID-19. *Eur J Hum Genetics* 28: 728–731. <https://doi.org/10.1038/s41431-020-0634-8>
 35. Tieri P, Nardini C, Dent JE (2015) Multi-omic data integration. *Front Cell Dev Biol* 3:46. <https://doi.org/10.3389/fcell.2015.00046>
 36. Alberich-Bayarri A, Neri E, Martí-Bonmatí L (2019) Imaging biomarkers and imaging biobanks. In: *Artificial Intelligence in Medical Imaging*. Springer, Cham, pp 119–126. https://doi.org/10.1007/978-3-319-94878-2_10

Publisher's Note

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

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)