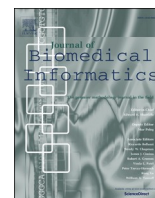
















Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Special Communication

Implementation of a Digital Maturity Framework for Biobanking[☆]

Federica Rossi^{a,1} , Davide Fragnito^{a,1} , Antonella Cruoglio^{a,1} , Ramona Palombo^a , Alice Massacci^a, Alessandro Sulis^b, Vittorio Meloni^b , Sara Casati^a, Antonella Mirabile^a, Andrea Manconi^c , Luciano Milanese^c , Gennaro Ciliberto^d, Monica Forni^e , Valentina Adami^f , Massimiliano Borsani^a , Claudia Miele^a, Marialuisa Lavitrano^{g,1} , Matteo Pallocca^{a,1,*} 

^a Istituto degli Endotipi in Oncologia, Metabolismo e Immunologia "G. Salvatore" (IEOMI), Naples, Italy^b CRS4, Center for Advanced Studies, Research and Development in Sardinia, 09050 PULA, CA, Italy^c Institute of Biochemical Technologies, National Research Council (CNR), 20054 Segrate, Milan, Italy^d Scientific Direction, IRCCS Regina Elena National Cancer Institute, 00144 Rome, Italy^e Department of Medical and Surgical Sciences (DIMEC), Interdepartmental Center for Industrial Research in Health Sciences and Technologies (CIRI-SDV), University of Bologna (UniBo), 40126 Bologna, Italy^f CIBIO - Department of Cellular, Computational and Integrative Biology, University of Trento 38123 Trento, Italy^g School of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy

ARTICLE INFO

Keywords:

Biobanking

Biobanks

Data Management

Digital Framework

FAIR Principles

BBMRI

Research Infrastructures

ABSTRACT

Objective: Digitalization is a pillar of reproducible research and a mandatory requirement for Research Infrastructures. Biobanks must ensure a fully engineered and digitalized process towards data FAIRification. To this aim, the first step is to assess the current level of digitalization using quantitative metrics, which is particularly challenging given the multi-faceted regulatory and logistical nature of biobanking.

Methods: We developed a Biobanking digital assessment maturity framework, BB4FAIR, comprising a survey divided into three macro areas, namely IT infrastructure, personnel, and data annotation richness. Furthermore, we implemented an automated R/Shiny system to analyse survey responses and generate visual data representations. We piloted the tool on 46 Italian biobanks that in 2023 had signed the partner charter with BBMRI. A scoring table facilitated the tiering of digital maturity, highlighting areas requiring corrective action.

Results: The assessment revealed significant heterogeneity across the three macro-areas of digitalization: almost half of the biobanks feature adequate IT infrastructure and personnel, and a smaller proportion have robust data annotation capabilities. Notably, most biobanks reported having a Biobank IT Management System (BIMS) or an alternative that serves their purposes, yet they still collect the consent to biobanking for future purposes in paper format; the digitalization of informed consent is generally lacking. These findings highlight the need for targeted improvements in Biobank digitalization to enhance overall data FAIRness.

Conclusion: The survey results underscore a pressing need for enhanced IT training and improved data annotation resources within the BBMRI.it. Corrective actions on many lacking features and desiderata are ongoing in the context of the #NextGenerationEu "Strengthening BBMRI.it" project.

Abbreviations: BBMRI.it, Biobanking and Biomolecular Resources Research Infrastructure of Italy; BIMS, Biobank Information Management System; BRCs, Biological Resource Centers; CDM, Common Data Models; EHR, Electronic Health Records; ELSI, Ethical, Legal, and Social Issues; FAIR, Findable, Accessible, Interoperable, and Reusable; HL7-FHIR, Health Level Seven International Fast Healthcare Interoperability Resources; ISO, International Organization for Standardization; IT, Information Technology; LIMS, Laboratory Information Management System; OMOP, Observational Medical Outcomes Partnership.

[☆] This article is part of a special issue entitled: 'Data Generation in Healthcare Environments' published in Journal of Biomedical Informatics.

* Corresponding author.

E-mail address: matteo.pallocca@cnr.it (M. Pallocca).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.jbi.2025.104842>

Received 30 October 2024; Received in revised form 18 April 2025; Accepted 27 April 2025

Available online 7 May 2025

1532-0464/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Biobanks are gaining widespread acceptance and interest within the biomedical research community [1,2] as defined by the official International Organisation for Standardisation (ISO) document (ISO 20387:2018) [3]. Biobanks serve as service structures that guarantee several processes, including the collection, preparation, testing, storage and distribution of the biological materials and related data, as well as the management of returning research results. Importantly, biobanks ensure access to personal and biomedical data necessary for the discovery of new and targetable disease characteristics in accordance with legal and regulatory requirements. These data may include health records, lifestyle, demographic, and genetic data, up to more complex readouts such as –omics and imaging files.

Thus, the true value of biobanks lies in enabling researchers to access in a controlled way biological and data resources stored in compliance with international quality standards, while ensuring the privacy and rights of the biobank participants. Alignment with the FAIR principles provides a powerful framework for improving the discoverability and usability of research data [4–6]. This requires efficient information technology (IT) support and standardised and interoperable clinical annotation [7]. However, the effective management of data repositories poses significant challenges and requires careful consideration of many factors, including infrastructure, equipment, funding and recruitment of highly specialised personnel [8–12].

Many Italian biobanks, biological resource centers (BRCs), and sample collections are associated with BBMRI.it (Biobanking and Biomolecular Resources Research Infrastructure of Italy) [13]. However, biobanks may have different backgrounds and objectives influencing their structure and relative management systems. Therefore, the analysis of the current state of the Italian BBMRI biobanks has revealed some discrepancies and differences between the infrastructures.

In the overall drive towards data FAIRification, many maturity models have been implemented and tested in the European context. The Italian research hospital network has been assessed through the 10-year Health Big Data project [14], while the OECL-supported DIGICORE network implemented a specific maturity model for Cancer Outcomes [15,16].

We hereby discuss the implementation and testing of BB4FAIR, a framework for Biobanking Digital Maturity. We piloted the framework with 46 Italian biobanks associated with BBMRI.it, assigning each to a digital maturity tier and identifying priority areas for improvement. Our long-term vision is to consolidate the data engineering processes in mature biobanks, support intermediate-tier biobanks with additional resources, and handle corrective action for those in the starting tiers. Moreover, BB4FAIR comprises an automatized R framework that enables automated feedback generation and visualization of results. This can be applied in different contexts with a particular interest in digital maturity assessment, such as in biobanks and other service and/or research provider organizations that collect, store, manage, and share data and biological samples.

Statement of Significance	
Problem	Assessing the digitalization status of Biobanks. Achieving Findable, Accessible, Interoperable, and Reusable (FAIR) data principles in biobanking.
What is Already Known	Digitalization is crucial for fair access and reproducible research and alignment with FAIR principles. Effective management requires IT support, standardized clinical annotations, and specialized staff. Biobanks in Italy are numerous and diverse in terms of collected material, governance and purpose.
What This Paper Adds	Introduces BB4FAIR, a framework for assessing digital maturity of biobanks with quantitative metrics and automated analysis. Pilots the BB4FAIR framework with 46 Italian biobanks included in the Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium (BBMRI-ERIC) and highlights key areas for improvement.

2. Materials and Methods

2.1. Infrastructure context

The objective of this study was to assess 46 qualified biobanks that had formally joined BBMRI.it through the signature of the BBMRI-ERIC partner charter as of July 2023.

BBMRI-ERIC is a distributed research infrastructure of biomolecular resources based on a hub-and-spoke architecture in which each member state participates through a national node that provides expertise and services, specifically in Information Technology, Quality, and ELSI.

2.2. Survey dimensions and context description

We designed a survey with 38 questions addressing various crucial aspects of biobanking data and metadata digitalization (Table S1, Fig. 1). The first part gathers information about Biobank, affiliation and respondents, followed by questions focused on three main digital clusters: IT infrastructure facilities, data annotation richness and dedicated IT personnel, labelled in short with, *IT infrastructure*, *data*, and *personnel*, respectively. To our knowledge, this survey represents the first evaluation focused on the basic digital resources available to biobanks. The information collected from the BBMRI.it biobanks will serve as a starting point for tracking the digital maturity progress over time.

2.3. Definition of the three main digital areas and questions clustering

The survey sought to investigate three main digital areas: IT infrastructure facilities, data annotation richness, and dedicated IT staff.

IT infrastructures: Informatic and technological resources are of pivotal importance to enable biobanks to allocate specific services such as Federated computing nodes and advanced, redundant, and disaster-ready data storage facilities. In the Italian context, a migration to the cloud of several critical healthcare services is still ongoing [17], but for Biobanks, a hybrid local cloud approach is still preferred to enable fast setup of GDPR-compliant clinical research services. These requirements are tightly linked to internal equipment since centres that also act as core facilities, producing terabytes of –omics data for internal and external purposes, necessitate ensuring additional layers of structured and secure data processes, especially when dealing with sensitive molecular signatures. Moreover, the presence of specific management software is accounted as infrastructure, specifically regarding Biobank Information Management System (BIMS) [18–20]. In the survey, we also requested information about data management, focusing on data annotation. This includes the possibility of connecting biobank data to other clinical data management systems or Electronic Health Records (EHR), the presence of collections shared with specialized networks, and the use of a digital-electronic informed consent. For this purpose, we evaluated questions 14, 16, 17, 19, 20 (Table S1 and Table S2).

Data annotation richness: *Data annotation richness* refers to all tools and procedures that make biobanked samples findable, including ontologies, and that help to define the magnitude of sub-collections thanks to structured data and annotations containing at least the minimal fields required by the BBMRI-ERIC Federated Platform [21]. Richer annotations regarding clinical phenotypes, longitudinal data, and linked-omics data enable to answer nuanced research questions without expensive data recollection and analysis. The questions referring to this aspect of digital maturity were 15, 22, 30, 34, and 36 (Table S1 and Table S2).

Dedicated personnel: In this area, we sought to evaluate the presence of *dedicated personnel* specialized in IT, Data Engineering, or Data Analysis. Experienced biobank staff members empower the Biobank to design data processes and speed up data cleaning and collection tasks. We also examined the presence of human resources dedicated

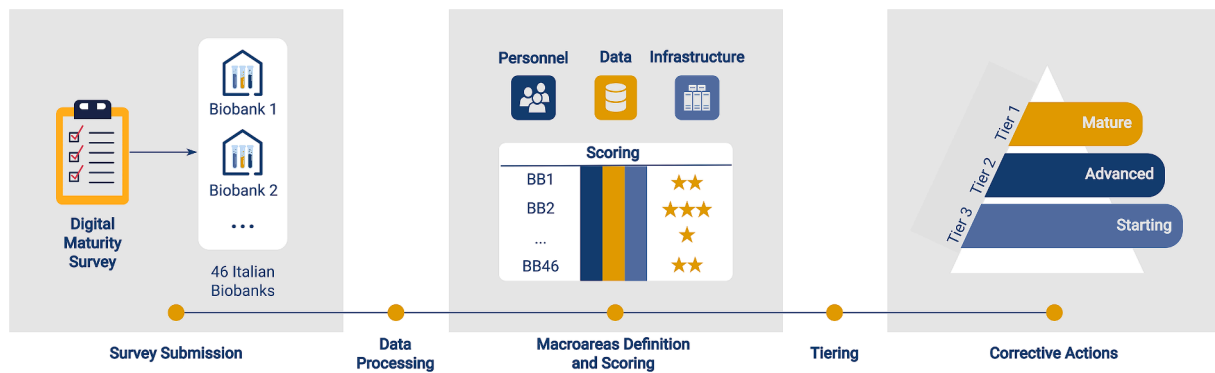


Fig. 1. Graphical representation of the Digital Maturity Framework, from Biobank surveying (left) to content clustering and analysis (middle) to tiering processes (right).

exclusively to biobank digitalization and what kind of professional skills they have, with a particular emphasis on expertise in data annotation and modelling and the usage of Common Data Models (CDM) such as OMOP and HL7-FHIR. Knowledge of Common Data Models is an advanced feature that has been classified in the personnel and not in the data section since the conversion of Biobank databases in standardized models requires a strong know-how of mixture semantics between ontologies, computer science, and database programming. The BBMRI.it is working towards a unified and open-source model for Extract, Transform and Load (ETL) processes on the minimal required dataset [22]. To this end, questions n. 2, 6, 8, and 10 of the survey were evaluated (Table S1 and Table S2). To evaluate the digitalization status of BBMRI.it biobanks, we created a scoring system based on the survey responses. First, we normalized and standardized the qualitative answers to enable automated scoring and visualization. This step required manual curation of the descriptive answers and data cleaning. Then, we developed a visualization script that generates plots for each question, allowing us to identify general trends within the biobank community.

To rank Biobanks by their digitalization status, we developed a scoring strategy based on a matrix that converts qualitative responses into quantitative scores. We defined the possible values for each response and used these to construct the scoring matrix. Each row of the matrix corresponds to a survey question and contains different scores based on the potential responses. Some questions gave a binary score based on whether a feature was present or absent. However, other features needed a different scoring scale, and in this case, scores were weighted according to the item's relevance to digitalization. For example, the usage of a proper BIMS has a higher score than the presence of a massive storage system (Table S2). In the cases in which the answers did not provide full compliance with the specific feature, we assigned a lower score defined as *partial response*. For instance, some Biobanks do not have a proper BIMS in place, but they do have intermediate software dealing with cryopreservation or other tools simply containing a relational database with annotation. That situation is regarded as a first step towards an engineered metadata process.

The scoring workflow considers both the answers file and scoring matrix, connects them by the question ID to which they refer and assigns a value according to the response. The script's output is a table with rows corresponding to biobanks and columns filled with answer scores, with the last column containing the summarized result.

For the limited set of non-binary features, we adjusted the score values to facilitate visualization of the results (Table S2). We scaled them to the unit by dividing by the higher value initially assigned to them, obtaining for all the answers the same lower and higher value, respectively 0 and 1.

2.4. Biobank Tiering System

Biobanks were divided into Tiers based on their scores, partitioning

by tertiles. Biobanks with the highest score in the top third are placed in the Mature Tier, those in the middle are classified as Advanced Tier, while the bottom third represents the Starting Tier. The R script processes the scoring module and assigns each Biobank to its corresponding tier. The whole process, from data normalization, plot visualization, to scoring and tiering, is implemented in an R script repository [23]. To ensure portability and reproducibility, the required packages have been embedded in a *Conda* repository list and virtualized in a Docker image [24,25]. The required input is a CSV/XLSX file where each row contains the responses from a single Biobank, and each column corresponds to a survey question. For each question, the script processes the corresponding column, performs string normalization to standardize the responses, counts the occurrences of each response and then generates a plot for results visualization.

3. Results

3.1. Feature analysis and Tiering results

46 Biobanks participated in the study. Of these, 37 gave explicit consent to publishing their results in a pseudonymised format. Exploratory analysis of all questions across the three dimensions for the biobanks that gave consent is depicted in Fig. 2, included in the public ShinyApp. All these results and subsequent comments must take into consideration the high heterogeneity in the BBMRI.it landscape, which contains disease-oriented, population-based, and multi-specialistic biobanks that exhibit slight differences in digitalization requirements (Supplementary Material)

When considering the density of all the answers, the score distribution is bimodal, reflecting the feature weights, which are mostly binary.

Furthermore, almost half of the biobanks meet the requisites for IT *infrastructure* and *personnel* (respectively $\sim 51\%$ (51.4) and 50% (50.7)), while the *data* cluster is enriched in only about 40% (38.9) of the sample population (Fig. 2A).

When considering the answers from a more fine-grained view, some features are transversally present or absent across the network which may easily drive infrastructure-wide corrective actions (Fig. 2B).

On the *data* landscape, all biobanks declare having at least one of the annotations defined at a macro level (e.g., clinical data, phenotype, genomics) (100%). The number of biobanks able to link samples to national registries is quite high (62.2%), while the refinement on specific clinical data or internal data warehouses is lower (respectively 16.2% and 30.5%). This result can be associated with the different mission and governance of the Italian infrastructures; for instance, population biobanks are not required to have a data warehouse, lacking the usual *data silos issue* present in disease-oriented biobanks. Finally, only the paper form was in use to trace and document the informed consent to biobanking for future purposes, de facto its digitalization is virtually absent in the biobank environment (94.6%).

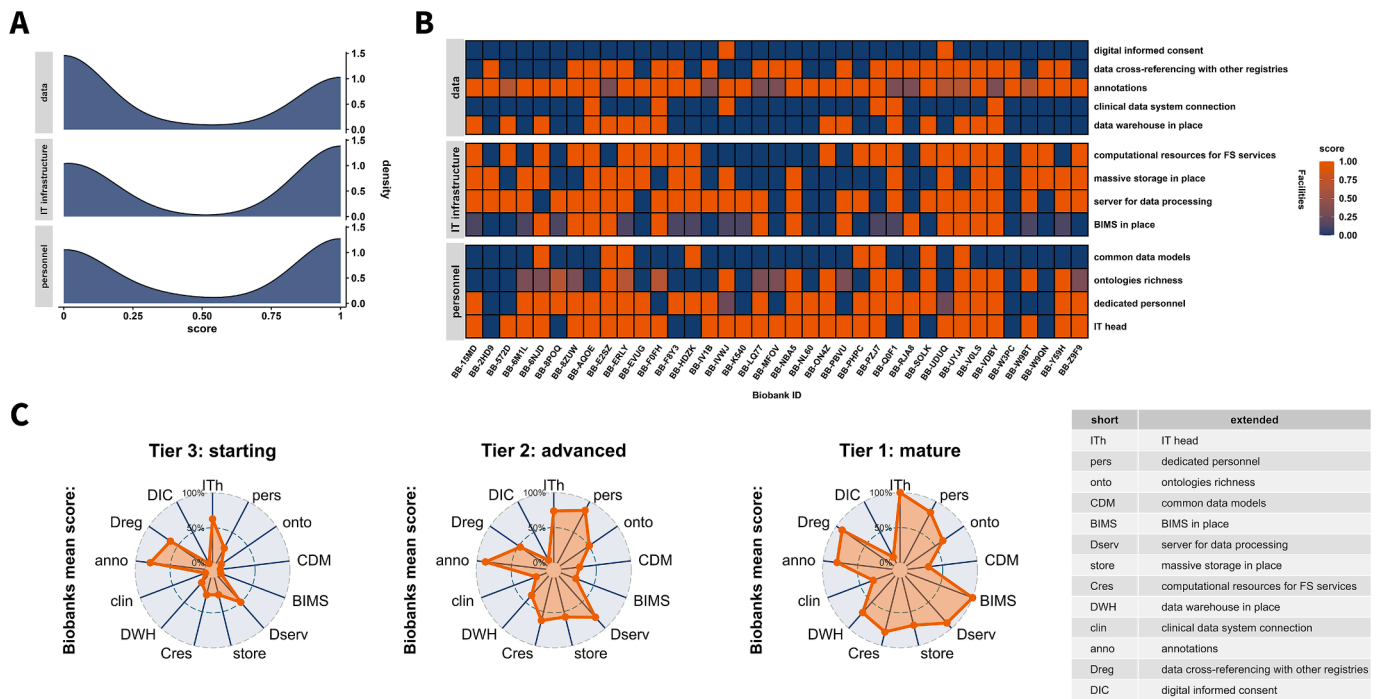


Fig. 2. density plot of the distribution of the BB4FAIR score in the three different macro areas (A); heatmap of the BB4FAIR score in the single facilities for the pseudonymized biobanks (B); radar plot of the mean BB4FAIR score in the different biobanks facilities split by three-tier class from starting to mature (C).

When considering the *IT infrastructures* associated with biobanks, most of them declare having access to an IT infrastructure able to process data, allocate resources for specific services such as Federated Search, and with more than 20 TB massive storage (75.7 %, 56.5 %, 54.1 %, respectively). More than half declare having a biobank management software (BIMS) (67.6 %), while the rest have an alternative that serves their purposes or plan to acquire one (32.4 %). In particular, when performing data cleaning steps in the BIMS-related responses, we concluded that one third of biobanks have a proper BIMS in place, another third employs relational database or other management LIMSeS, and another third still need to acquire any system or rely on other department software (e.g., the Pathology department) (Fig. S2).

Regarding the personnel-related questions, almost all biobanks are empowered with an IT coordinator and dedicated staff to data management (78.4 % and 70.3 %, respectively). The nature of this figure varies in specialization, from Data Stewards or Data Scientists to more institutional profiles, such as the head of IT services in cases of smaller institutions. We did not further investigate this aspect, which is inherently heterogeneous along with the biobank landscape. On the other hand, the know-how regarding Common Data Models is very low (21.2 %), as witnessed in many European endeavours aimed at improving infrastructure interoperability [26–28]. Finally, the *ontology richness* feature describes that more than half biobanks employ standardized nomenclatures in routine annotation (59.5 %).

To make infrastructure-wise considerations, we produced specific visualizations for all scores in the three maturity tiers, helpful to personalize node support (Fig. 2C). Starting Tier biobanks are characterized by the availability of registry and clinical data without a BIMS system in place. In contrast, advanced Tier Biobanks do exhibit more infrastructure features and dedicated personnel. Finally, the transition from tiers 2 and 3 to the Mature Tier is characterized by the acquisition of a BIMS and additional IT infrastructure scores such as data service access and storage (Fig. 2C). Taken together, all these results point to the main factors to be dealt with to improve Biobank’s digitalization and overall data FAIRness.

3.2. Results Dissemination

Gathering all the data made it necessary to disclose the scores each biobank obtained and make them available in an aggregated and pseudonymised manner. To complete this process, a ShinyApp page was developed and made publicly available following the URL: https://bb4fair.taeng.shinyapps.io/bb4fair_app/.

The BB4FAIR tool is useful for viewing the survey results both in individual and aggregate form. In the main web interface, the four panel’s tabs in the left-hand menu (Fig. 3A) allow to navigate through the entire survey and grasp its various nuances. The *scores* section (Fig. 3B) summarises the main characteristics of the biobanks and presents the data in pseudonymised form, with some variables scaled down in relation to the total score to ensure a homogeneous weighting across the different institutions. The *feature analysis* section (Fig. 3C) shows the data in percentage and absolute form for individual responses to the survey questions. The *visualization* section (Fig. 3D) gives an aggregate graphical representation of the entire survey, showing the frequency density and a heatmap with the maturity percentage in each facility. Finally, in the *tiering* section (Fig. 3E), an overview of the scores for the single biobank is reported, highlighting the mean trend of the biobanks present in the national node at the time the questionnaire was administered, and a ranking based on the division into tertiles of the sum of the total scores obtained.

4. Discussion

Data FAIRness is a requirement designed by all research stakeholders to improve research quality and reproducibility. Implementing maturity frameworks is a strategy that sets specific goals for RI networks. Importantly, these characteristics represent the basic foundation for the evolving biobanking sector. The quality of biological samples is guaranteed by timely technical verification, but the associated data quality is yet to be defined. As a minimum requirement for a FAIRification process, data must be collected, stored, and shared correctly. These processes should be performed by dedicated personnel who must be trained and informed about the standards for data elements, data types, and file

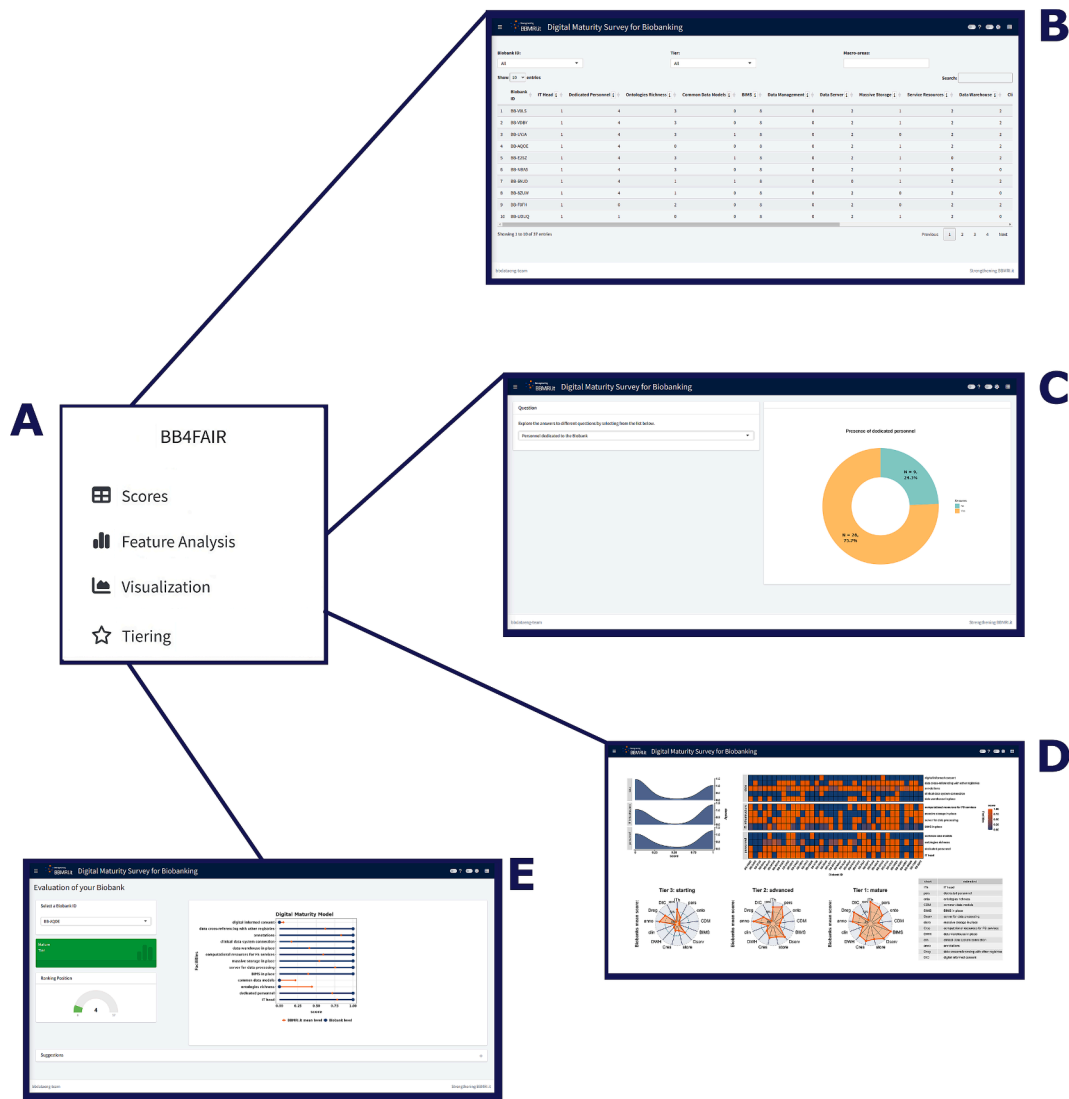


Fig. 3. Overview of the BB4FAIR Shiny app sections. (A) Main menu with the four main sections: Scores, Feature Analysis, Visualization, and Tiering. (B) The Scores panel displays pseudonymized score data for each biobank. (C) Feature Analysis panel showing the distribution of responses to individual questions. (D) Visualization panel presenting an aggregate graphical representation of responses across macro-areas, facilities, and tiers. (E) Tiering panel illustrating individual biobank performance over the national average and ranking based on tertiles of total scores.

formats.

Importantly, these results confirm the strategy that a few aspects must be improved at the national level: these aims are pillars of the “*Strengthening BBMRI.it*” project [29]. Among these, the knowledge and usage of Common Data Models [22] and the implementation of a digitalized informed consent based on the Common Condition of Use Elements aimed at characterizing the samples biobanked with the participants’ preferences as condition of use [30].

Having IT-savvy dedicated personnel is a natural desideratum of smaller nodes that cannot rely on larger centralized IT facilities. Unfortunately, the job market is pressing the need for data experts of varying complexity, from data stewards to AI engineers, that low-income and public nodes struggle to retain [31]. This situation is pushing national nodes like BBMRI.it into recruiting and training distributed teams of Data Engineers dedicated to biomedical data management.

This work aimed at creating a gold-standard model for biobanks’ digital maturity that complies with the FAIR principles and measures how digital maturity progresses over time, rather than serve as a method of classification of biobanks.

Importantly, this digital assessment could be extended to all BBMRI-ERIC biobanks. The common objectives and principles are specified in the Partner Charter and signed by all members, creating an environment driven by common goals and challenges.

There is also the possibility of applying this model to other types of healthcare-providing institutions involved in research. Our approach is based on implementations of digital tools that can greatly facilitate scientific research by helping to make the data and results generated more reproducible, accessible and managed in accordance with current regulations.

Finally, this and other similar efforts can pave the way for integrating biobanks in Health Information Systems since the improvement of interoperability processes is proceeding in parallel and will finally converge, from where data and metadata are collected towards biobank data storage. We consider this vision particularly important for the Italian EHR system, which can be considered a pilot for other European systems facilitating compliance with European Health Data Space requirements [32].

Digital maturity correlates with several Key Performance Indicators (KPIs), such as the number of accesses and FAIRified collections. In

addition, it can be used to evaluate the progress against a performance baseline.

Finally, this framework could be aligned with existing general frameworks like the FAIR toolkit implemented by the Pistoia Alliance [33,34], with a putative mapping model from our maturity levels to standardized FAIR levels. Nonetheless we consider this a valuable and scalable tool built specifically for the biobanking domain.

5. Conclusions

The discussed digital framework was tested to assess the status of BBMRI Italian biobanks at a specific time. Biobanks are rapidly evolving towards more complex data infrastructures. It is expected that this tool, which is easy to manage and update, will be able to provide a concise but qualified picture of the progress of each biobank over time, with the possibility of updating in certain period its progress and guiding the corrective actions among the facilities in the different macro-areas for the entire node.

6. Funding Sources

This publication has been supported by the funding from Italy's National Recovery and Resilience Plan (NRRP), supported by the European Union – NextGenerationEU, Mission 4 “Education and Research”, Component 2 “From Research to Business”, Investment 3.1 “Fund for the realisation of an integrated system of research and innovation infrastructures”, project code IR0000031 - *Strengthening BBMRI.it* - CUP B53C22001820006, and partly supported from the fund “The Italian Network of Biobanks and Biomolecular Resources (BBMRI-it)” FOE 2023 granted to ML/UNIMIB.

CRedit authorship contribution statement

Federica Rossi: Writing – original draft, Visualization, Software, Formal analysis. **Davide Fragnito:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Antonella Cruoglio:** Visualization, Software, Data curation, Conceptualization. **Ramona Palombo:** Writing – original draft, Methodology, Formal analysis. **Alice Massacci:** Writing – review & editing, Conceptualization. **Alessandro Sulis:** Resources, Investigation. **Vittorio Meloni:** Resources, Investigation. **Sara Casati:** Resources, Investigation. **Antonella Mirabile:** Resources, Investigation. **Andrea Manconi:** Resources, Investigation. **Luciano Milanese:** Resources, Investigation. **Gennaro Ciliberto:** Resources, Investigation. **Monica Forni:** Writing – review & editing. **Valentina Adami:** Writing – review & editing. **Massimiliano Borsani:** Writing – review & editing. **Claudia Miele:** Writing – review & editing. **Marialuisa Lavitrano:** Supervision, Resources, Investigation, Funding acquisition. **Matteo Pallocca:** Writing – original draft, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all the BBMRI Italian biobanks for participating in the survey and providing valuable feedback. We would like to especially thank for their collaboration the Biobank Heads and the personnel who participated in the survey, specifically Rita Lawlor from ARC-Net, Deborah Novelli from Banca Biologica di Ricerca Cardiovascolare Istituto di Ricerche Farmacologiche Mario Negri IRCCS - SATURNE, Andrea Zangrando from Banca Biologica Oncologica Pediatrica, Giuseppina Di Stefano and Anna Rita Bonfigli from Biobanca Istituzionale per la

Ricerca sull'Invecchiamento - BioGer, Elena Pegoraro and Marco Agostini from Banca Neuromuscolare di Tessuti e DNA, Giusy Esposito from BCU Imaging Biobank, Rosanna Cardani from COVID-19 Biocor, Alessia Di Sapio from Biobanca del centro di Riferimento Regionale per la Sclerosi Multipla - CRESM, Domenico Coviello and Chiara Baldo from Biobanca del Laboratorio di Genetica Umana, Simonetta Bisanzì from Biobanca del Laboratorio Regionale Prevenzione Oncologica di ISPRO, Simonetta Piana and Eleonora Zanetti from Biobanca di ricerca AUSL di Reggio Emilia - IRCCS, Tiziana Franchin from Biobanca Di Ricerca OPBG, Ornella Parolini from Biobanca di Ricerca per la Medicina Personalizzata, Luisa Benussi from Biobanca IRCCS Centro San Giovanni Di Dio Fatebenefratelli, Gennaro Ciliberto, Simona Di Martino and Giovanni Cigliana from Biobanca IRCCS Istituto Nazionale Tumori Regina Elena - BBIRE, Alfredo Zito from Biobanca Istituzionale Bari - BBI Bari, Alfredo Budillon from Biobanca Istituzionale - BBIIstituzionale, Giovanni Apolone and Laura Cattaneo from Biobanca Istituzionale INT - Istituto Tumori Milano, Fabio Ruggero Maria Corsi from Biobanca per la ricerca Bruno Boerci, Luca Valenti from Biobanca 'POLI-MI', Elda Judica and Massimo Corbo from CCPP Biobank, Luca Sangiorgi from Centro di Risorse Biologiche Istituto Ortopedico Rizzoli - CRB IOR, Roberta Libener from Banca Biologica del mesotelio maligno - Alessandria Biobank, Cristina Tresoldi from Centro Risorse Biologiche - Ospedale San Raffaele, Carla Fontana and Stefania Carrara from Banca Biologica INMI L. Spallanzani, Alessandro Degrandi from Cooperative Health Research in South Tyrol Study Biobank, Paola Nincheri from da Vinci European BioBank, Alessandra Renieri and Chiara Fallerini from Genetic and COVID-19 Biobank of Siena, Daniela Pistillo from Humanitas Hospital Biological Resources Center, Maria Beatrice Boniotti from IZSLER Biobanking of Veterinary Resources, Maria Benedetta Donati from Centro Risorse Biologiche Neuromed Biobanking Center, Sara Gilbertini from Neuromuscular Disease Biobank, Chiara Piubelli from Tropica Biobank - IRCCS Ospedale Sacro Cuore Don Calabria, Davide Guarneri from UOSD Biobanca - ASST Papa Giovanni XXIII, Daniela Capello from UPO Biobank, Salvatore Pucciarelli from Chirurgia Generale 3 Tumour Tissue Biobank - Biobanca Università Padova, Valentina Ancarani from Centro Risorse Biologiche IRST.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jbi.2025.104842>.

References

- [1] P.H.J. Riegman, M.M. Morente, F. Betsou, P. de Blasio, P. Geary, Biobanking for better healthcare, *Mol Oncol* 2 (2008) 213–222, <https://doi.org/10.1016/j.molonc.2008.07.004>.
- [2] L. Annaratone, G. De Palma, G. Bonizzi, A. Sapino, G. Botti, E. Berrino, C. Mannelli, P. Arcella, S. Di Martino, A. Steffan, M.G. Daidone, V. Canzonieri, B. Parodi, A.V. Paradiso, M. Barberis, C. Marchiò, O. behalf of A.C. il C. (ACC) P. and B.W. Group, Basic principles of biobanking: from biological samples to precision medicine for patients, *Virchows Archiv* 479 (2021) 233–246, <https://doi.org/10.1007/s00428-021-03151-0>.
- [3] Iso, accessed August 2, - Biotechnology — Biobanking — General Requirements for Biobanking, (n.d.), h 20387 (2024) 2018, <https://www.iso.org/standard/67888.html>.
- [4] K. Im D. Gui W.H. Yong An Introduction to Hardware, Software, and Other Information Technology Needs of Biomedical Biobanks W.H. Yong *Biobanking: Methods and Protocols*, Springer 2019 New York, New York, NY 17 29 10.1007/978-1-4939-8935-5_3.
- [5] M.D. Wilkinson, M. Dumontier, J. Ij, G. Aalbersberg, M. Appleton, A. Axton, N. Baak, J.W. Blomberg, L.B. da Boiten, P.E. Silva Santos, J. Bourne, A. Bouwman, T. Brookes, M. Clark, I. Crosas, O. Dillo, S. Dumon, C.T. Edmunds, R. Evelo, A. Finkers, A.J.G. Gonzalez-Beltran, P. Gray, C. Groth, J.S. Goble, J. Grethe, P.A.C. Heringa, T. Hoen, R. Hoof, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data* 3 (2016), <https://doi.org/10.1038/SDATA.2016.18>.

- [6] FAIR Cookbook, (n.d.). <https://faircookbook.elixir-europe.org/content/home.html> (accessed August 2, 2024).
- [7] Hu., B. Prokosch A., Ganslandt T., Hummel M., Kiehnopf M., Sax U., Ückert F., Semler S., IT Infrastructure Components for Biobanking, *Appl Clin, Inform* 01 (2010) 419–429, <https://doi.org/10.4338/ACI-2010-05-RA-0034>.
- [8] G. Bonizzi, L. Zattoni, M. Capra, C. Cassi, G. Taliento, M. Ivanova, E. Guerini-Rocco, M. Fumagalli, M. Monturano, A. Albini, G. Viale, R. Orecchia, N. Fusco, Standard operating procedures for biobank in oncology, *Front Mol Biosci* 9 (2022) 967310, <https://doi.org/10.3389/fmolb.2022.967310>.
- [9] R. Asimwe, S. Lam, S. Leung, S. Wang, R. Wan, A. Tinker, J.N. McAlpine, M.M. M. Woo, D.G. Huntsman, A. Talhouk, From biobank and data silos into a data commons: convergence to support translational medicine, *J Transl Med* 19 (2021) 1–13, <https://doi.org/10.1177/02611929221113976>.
- [10] A. Chróścicka, A. Paluch, I. Kalaszczynska, I. Szablowska-Gadomska, Biobank Personnel — The Key to its Success, *Altern. Lab. Anim.* 50 (2022) 275–281, <https://doi.org/10.1177/02611929221113976>.
- [11] R.R. Williams, D. Gupta, W.H. Yong, Orientation and Training of New Biobank Personnel, *Methods Mol Biol* 1897 (2019) 51, https://doi.org/10.1007/978-1-4939-8935-5_6.
- [12] C.B. Albiñana, M. Pallocca, H. Fenton, W. Sopwith, C. Van Eden, O. Akre, A. Auranen, F. Bocquet, M. Borges, E. Calvo, J. Corkett, S. Di Cosimo, N. Gentili, J. Guérin, S. Jor, T. Kazda, A. Kolar, T. Kuschel, M.J. Lostes, C. Paratore, P. Pedrazzoli, M. Petrovic, J. Raid, M. Roche, C. Schatz, J. Thonnard, G. Tonon, A. Traverso, A. Wolf, A.H. Zedan, P. Mahon, Developing PRISM: A Pragmatic Institutional Survey and Bench Marking Tool to Measure Digital Research Maturity of Cancer Centers, *Appl Clin, Inform* 15 (2024) 743–750, <https://doi.org/10.1055/S-0044-1788331/ID/OR202310RA0231-8/BIB>.
- [13] BBMRI.it, (n.d.). <https://www.bbMRI.it/> (accessed July 29, 2024).
- [14] Health Big Data - Alleanza contro il cancro, (n.d.). <https://alleanzacontroilcancro.it/en/progetti/italia/health-big-data/> (accessed July 31, 2024).
- [15] Carlos Berenguer Albiñana et. al, PRISM for Cancer Outcomes: PRagmatic Institutional Survey and benchMarking for Cancer Outcome Research, *Appl Clin Inform* (2024).
- [16] P.F. Zhang, X.H. Zheng, X.Z. Li, L. Sun, W.H. Jia, Informatics Management of Tumor Specimens in the Era of Big Data, Challenges and Solutions, <https://home.liebertpub.com/bio> 19 (2021) 531–542, <https://doi.org/10.1089/BIO.2020.0084>.
- [17] Polo Strategico Nazionale: il cloud sicuro per l'Italia digitale, (n.d.). <https://www.polostrategiconazionale.it/> (accessed July 31, 2024).
- [18] S. Paul, A. Gade, S. Mallipeddi, The State of Cloud-Based Biospecimen and Biobank Data Management Tools, <https://home.liebertpub.com/bio> 15 (2017) 169–172, <https://doi.org/10.1089/BIO.2017.0019>.
- [19] Engels Căcilia, D.P. Brucker, M. Neumann, How to choose your perfect Biobank Information Management System, *Zenodo* (2023). <https://zenodo.org/records/10063122> (accessed July 31, 2024).
- [20] L. Jacotot, M. Woodward, A. de Montalier, P. Vaglio, Utilizing Modular Biobanking Software in Different Types of Biobanking Activities, *Biopreserv Biobank* 20 (2022) 417–422, <https://doi.org/10.1089/bio.2022.0076>.
- [21] Join the BBMRI-ERIC Federated Platform - BBMRI-ERIC, (n.d.). <https://www.bbMRI-ERIC.eu/federated-platform/> (accessed August 2, 2024).
- [22] Antonella Cruoglio et. al, An Extract, Transform, Load foundation for Biobank Data Interoperability: from a-small-FIRE, . <https://zenodo.org/records/14012269>.
- [23] bbdataeng/BB4FAIR: Implementation of a Digital Maturity Framework for Biobanking., (n.d.). <https://github.com/bbdataeng/BB4FAIR> (accessed July 31, 2024).
- [24] conda.org | conda.org, (n.d.). <https://conda.org/> (accessed July 31, 2024).
- [25] Docker: Lightweight Linux Containers for Consistent Development and Deployment | Linux Journal, (n.d.). <https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment> (accessed July 31, 2024).
- [26] EvolveBBMRI - BBMRI-ERIC, (n.d.). <https://www.bbMRI-ERIC.eu/scientific-collaboration/evolvebbMRI/> (accessed August 2, 2024).
- [27] DigiCore - Homepage, (n.d.). <https://digicore-cancer.eu/> (accessed August 2, 2024).
- [28] EHDEN | IMI Innovative Medicines Initiative, (n.d.). <https://www.imi.europa.eu/projects-results/project-factsheets/ehden> (accessed August 2, 2024).
- [29] M. Borsani, M. Forni, S. Casati, M. Pallocca, A. Manconi, A. Scarpa, V. Adami, G. Falco, M. De Felice, M. Lavitrano, C. Miele, Strengthening the Italian National node of BBMRI, *BBMRI.it - ORAL PRESENTATION*, in: *EUROPE BIOBANK WEEK*, in: *European Bibank Week Programme Committee and BBMRI-ERIC's Department of Outreach, Education and Communication*, Wien, 2024, pp. 43–44.
- [30] S. Casati, A. Mirabile, M. Pallocca, S. Mellone, M. Lavitrano, Digital informed consent/assent matrix based on Common Condition of Use Elements: Coproducing an APP to track the process, return results, re-contact - *ORAL PRESENTATION*, in: *EUROPE BIOBANK WEEK*, in: *European Bibank Week Programme Committee and BBMRI-ERIC's Department of Outreach, Education and Communication*, Wien, 2024, p. 49.
- [31] S. Gaudin Expert tips for hiring (and retaining) data scientists 2022.
- [32] The Italian health data system is broken, *Lancet Reg. Health Eur.* 48 (2025) 101206, <https://doi.org/10.1016/j.lanep.2024.101206>.
- [33] Pistoia Alliance, The FAIR Toolkit for Life Science Industry, (n.d.). <https://fairtoolkit.pistoiaalliance.org/>.
- [34] D. Welter, N. Juty, P. Rocca-Serra, F. Xu, D. Henderson, W. Gu, J. Strubel, R. T. Giessmann, I. Emam, Y. Gadiya, T. Abbassi-Daloi, E. Alharbi, A.J.G. Gray, M. Courtot, P. Gribbon, V. Ioannidis, D.S. Reilly, N. Lynch, J.-W. Boiten, V. Satagopam, C. Goble, S.-A. Sansone, T. Burdett, FAIR in action - a flexible framework to guide FAIRification, *Sci Data* 10 (2023) 291, <https://doi.org/10.1038/s41597-023-02167-2>.