MEDICAL RESEARCH AGENCY

REGIONAL DIGITAL MEDICINE CENTRES

STANDARD

WARSAW, MARCH 2023

TABLE OF CONTENTS

1. THE CONCEPT OF REGIONAL DIGITAL MEDICINE CENTRES

Regional Digital Medicine Centres (RDMCs) are envisaged as structures that connect the databases of universities, teaching hospitals, biobanking facilities and other entities.

Digital Medicine Centres will serve as organisational units within the Applicant's internal structure responsible for the processes associated with the implementation of innovative digitisation instruments.

The objective in establishing Digital Medicine Centres is to standardise the collection and processing of high quality medical data for the purposes of research and analysis, as well as to ensure safe sharing of structured information.

Digital Medicine Centres are meant to organise the medical data management system at the principal universities and clinical institutions, and in the long term – create links between them and collect data from entities cooperating in the development of the RDMC. Depending on the source, the data obtained by the RDMC must be anonymised, pseudonymised or transparent. This will be decided by the entity that establishes the RDMC. In the present legal environment, an RDMC must be able to ensure full anonymisation of data in order to share them with other RDMCs. Data collected as part of the RDMC must not be used to feed private databases. The entities forming RDMCs undertake to share data among themselves to advance science and foster cooperation. When developing an RDMC, an entity must give due consideration to the changing legal environment and be prepared to share data with other entities.

When planning the establishment of an RDMC, an entity must carry out a number of analyses in order to ensure correct implementation of RDMC system functions in its structure. At the first stage, it is necessary, among other things, to carry out a legal, process, implementation, IT and technical analysis of the resources at hand and consider their potential for integration and interoperability. The project can be developed by gradual adoption of technology, known as agile iterative approach.

RDMCs will develop their own data handling tools and techniques, advancing the development of digital medicine.

The Centres will support clinical research at different stages (e.g. power calculations, results analysis).

Sharing and co-analysing findings by the Centres will bring about a breakthrough e.g. in rare diseases whose research is often limited to small groups of patients.

RDMCs will also transfer data as requested by the MRA's Digital Medicine Head Office to enable big data processing for the purposes of genomic analysis of the Polish population, in accordance with the EU programme of digital transformation of health[1] (Regulation (EU) 2021/522 of the European Parliament and of the Council of 24 March 2021 establishing a Programme for the Union's action in the field of health ('EU4Health Programme') for the period 2021-2027, and repealing Regulation (EU) No 282/2014).

2. RATIONALE FOR THE ESTABLISHMENT OF REGIONAL DIGITAL MEDICINE CENTRES

The Medical Research Agency has financed the establishment of Clinical Trials Support Centres as part of the Polish Clinical Trials Network. The next stage is for the entities that have CTSCs incorporated within their organisation to establish infrastructure supporting the development of medical digitisation and maximum use of health data of patients and other individuals who wish to contribute to a large scale programme for the development of digital medicine.

Regional Digital Medicine Centres (RDMCs) forming the Digital Medicine Centres Network (DMCN) supervised by the MRA – ultimately the MRA's Digital Medicine Head Office (DMHO) – will provide digital solutions to support real-time data analysis, clinical trials, and inpatient care, as well as retrospective analysis. The type of data collected will include e.g. data concerning health (within the meaning of Article 4(15) GDPR), omic data from samples collected by research institutions as part of their research projects (including non-commercial clinical trials financed by the MRA) described in the handbook *Quality Standards for Polish Biobanks, v.2.0*,[2] and other data collected by hospital information systems (HISs) or other repositories (such as imaging results, procedure reports and medical notes).

The data will be collected in accordance with the applicable quality standards (adapted by the entity in conformity with the prevailing European policy in the given area), ensuring their

[1] https://www.gov.pl/web/zdrowie/programy-unii-w-dziedzinie-zdrowia
[2] https://wydawnictwo.umw.edu.pl/upload/files/standardy_jakosci_dla_biobankow_polskich_2.0%281%29.pdf

analysable quality, secure sharing and storage. Further on, the MRA will work to make the collected data usable to third parties (governmental, academic) in keeping with the international standards, data security regulations and new EU legislature. The idea behind the Call follows the EU initiative for the deployment of a common European data space that provides access to large amounts of reliable data, at the same time ensuring control over the sources generating the data. Ultimately, European data spaces are to be created with the use of the European Open Science Cloud (EOSC), which is to provide European researchers and enterprises with access to an open multidisciplinary environment where they can search and reuse data.

Secondary data analysis is still a considerable challenge in logistic and legal terms, but it can provide a breakthrough in assessing diagnostic and therapeutic solutions both at the level of one Regional Digital Medicine Centre with support from other Centres of the Network, and – ultimately – for cumulative analysis by the Digital Medicine Head Office at the MRA. Creating a system that will enable maximum utilisation of medical data is therefore a reasonable and promising endeavour with a potential to transform the way medical data are processed in Poland.

3. OBJECTIVES OF REGIONAL DIGITAL MEDICINE CENTRES

The main objective of creating Regional Digital Medicine Centres within units that already have a CTSC and belong to the Polish Clinical Trials Network is to have the infrastructure, staff and systemic resources necessary to establish the Digital Medicine Centres Network. The placement of the RDMC within the organisational structure of an entity depends on the goals of the RDMC, its internal organisation, and the discretion of the parent entity.

The main objective will be supported by implementing the following ancillary objectives:

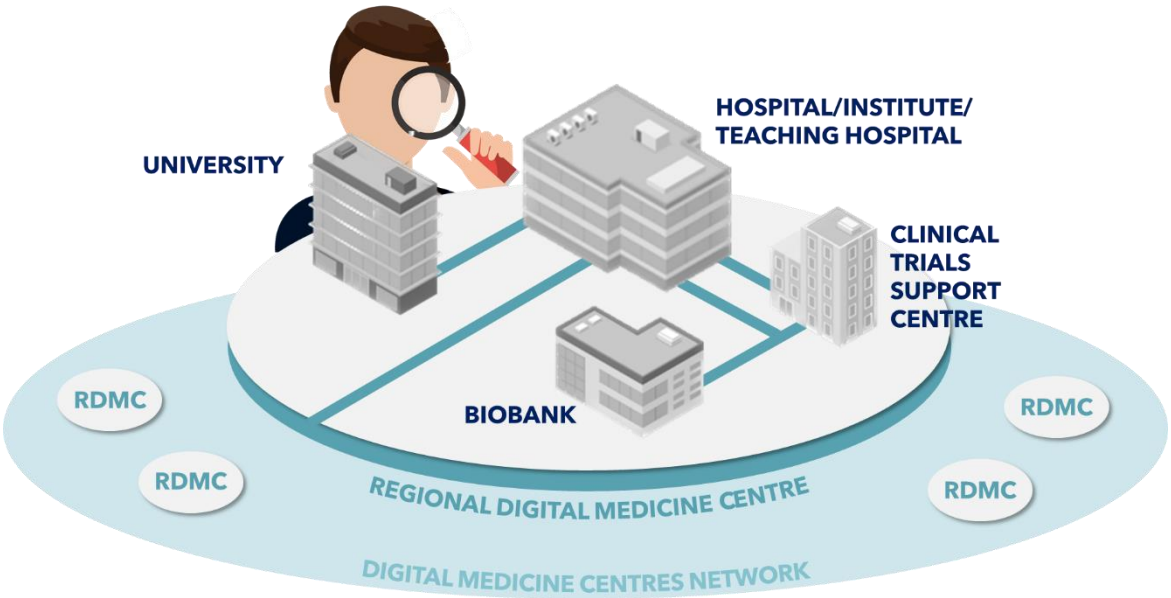
a) Assisting RDMCs in designing investigator-selected/indicated aspects of clinical trials (e.g. statistical methodologies), storing data and analysing them in real time in accordance with the existing standards, as well as undertaking retrospective data analysis, data anonymisation or pseudonymisation (if applicable), and final transfer in close cooperation with the MRA.

b) RDMCs generating high quality clinical and omic data sets that will provide the foundation for creating digital tools in the form of prognostic, predictive and therapeutic artificial intelligence algorithms to be used for the prevention, diagnostics and treatment of diseases.

c) RDMCs developing technical standards for sharing data and secondary data analysis in agreement with the MRA.
d) RDMCs producing the necessary algorithms and tools for the analysis of integrated data and creating efficient data security and data sharing software.
e) RDMCs creating the necessary facilities for direct data analysis (including omic data analysis) on data derived from biological samples collected from clinical trial subjects or other individuals who wish to contribute to the development of digital medicine.
f) Providing digital medicine services as part of research projects, particularly those financed by the MRA, e.g. as regards genetic testing results analysis, trial design in terms of digitisation or data analysis.

3.1. Responsibilities of the RDMCs

Regional Digital Medicine Centres will serve as organisational units within the parent entity's internal structure responsible for the processes associated with the implementation of innovative digitisation instruments. An RDMC may be established within the organisational structure of a university, hospital, institute, teaching hospital, CTSC or biobanking facility. The RDMC is obliged to provide support to other entities.



Examples of RDMCs' responsibilities:

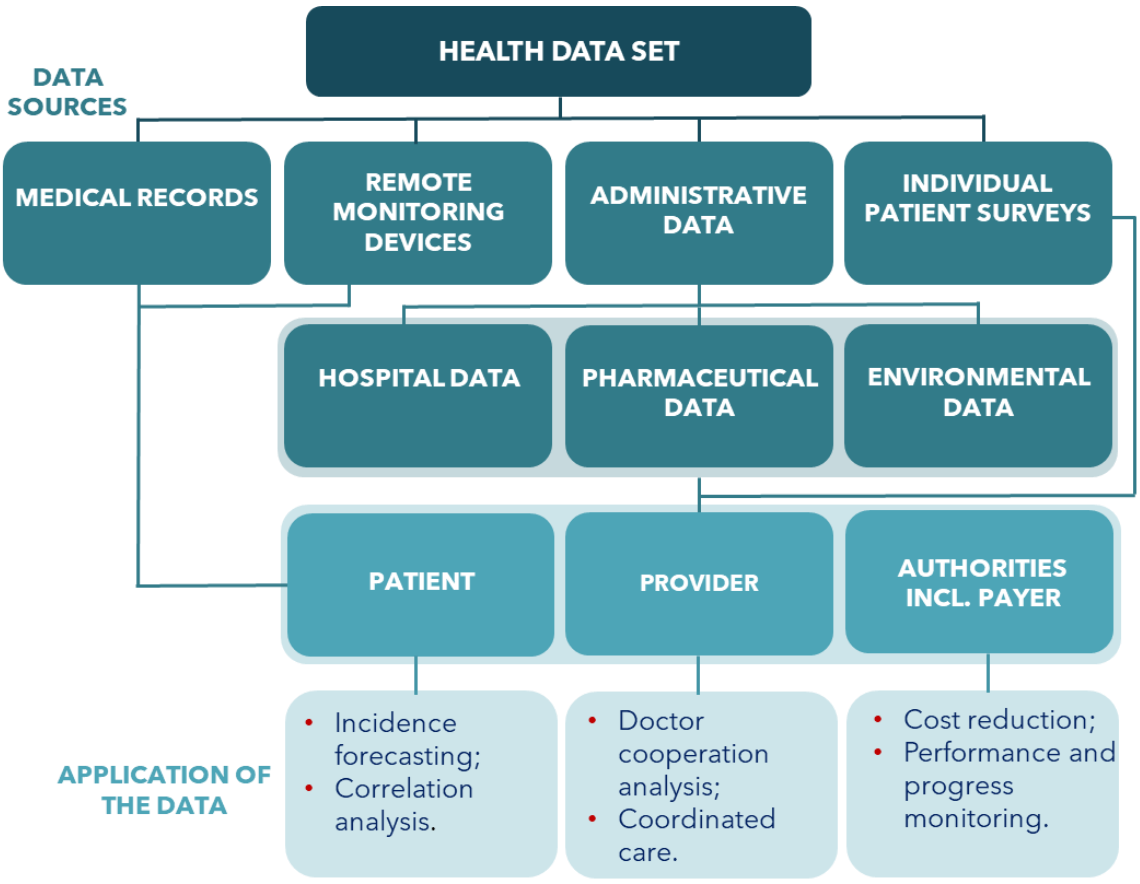
1. Creating, developing, promoting and implementing smart solutions, including artificial intelligence, e.g.:
    - creating digital tools in the form of prognostic, predictive and therapeutic artificial intelligence algorithms on the basis of clinical and omic data sets generated by the institution,
    - drug dosage algorithms,
    - algorithms generating adverse effect alerts when prescribing multidrug combinations,
    - algorithms supporting patient monitoring,
    - algorithms supporting treatment process management,
    - data integration and analysis as part of Picture Archiving and Communication Systems (PACSs) and Vendor Neutral Archives (VNAs) for archiving and sharing medical images from different sources,
    - data integration and analysis as part of Computer-Assisted Diagnosis (CAD) systems to facilitate diagnosis based on imaging,
    - treatment decision-making support systems.
    - Developing internal IT systems to allow for digitisation of medical records, creating AI algorithms supporting internal management processes.
    - Support of research and development associated with digital medicine.
    - Facilitating cooperation among doctors, IT specialists and biostatisticians in order to develop the AI algorithm.
    - Creating local databases that can be easily incorporated into the systemic solutions being established.
    - Supporting efforts to establish and develop standardised biobanking facilities. With a view to improving the quality of the data collected, biobanking facilities cooperating with Clinical Trials Support Centres are also to be created, developed and/or further equipped so that the samples stored and the data originating from them meet the necessary standards for efficient sharing of information and analyses within the DMC Network.

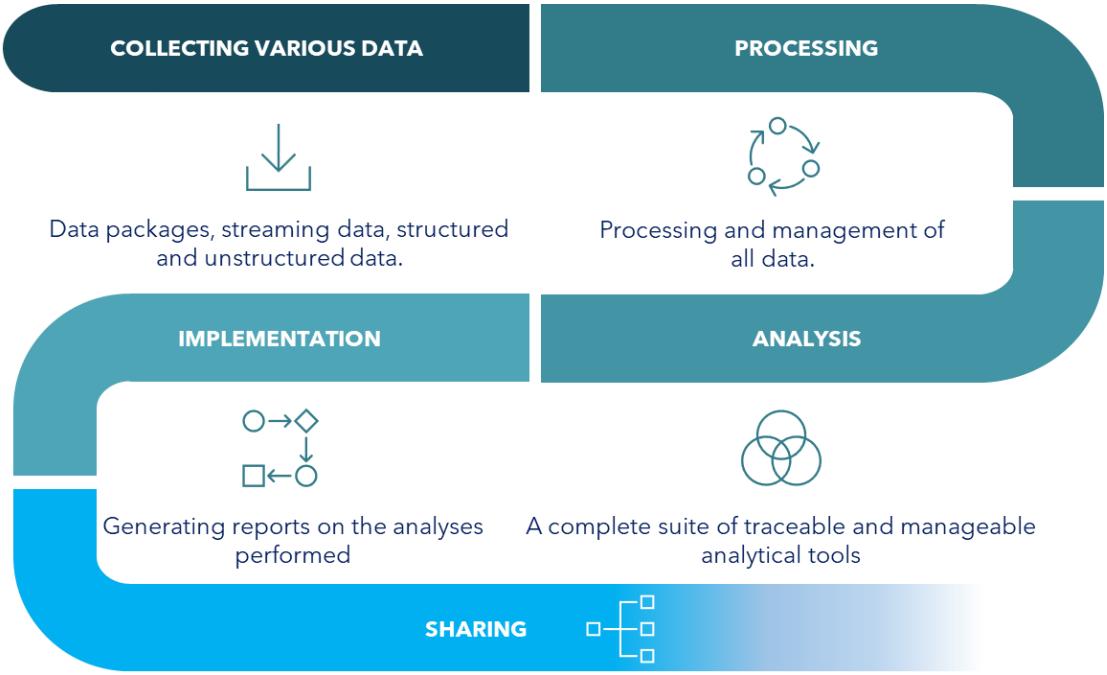
4. DATA SOURCES

4.1. Regional level (RDMCs at CTSCs)

The entity will need to present a detailed division of responsibilities and competence between the RDMC and CTSC, as well as methods of data flow between data sources and the risks involved.

The RDMC will be responsible for the management of medical data which include e.g. descriptive data, pharmaceutical information, imaging results, and test results from HIS or eCFR systems as well as other sources, such as patient monitoring devices or analysers. It is therefore important for the RDMC team to be familiar with the research process employed at the CTSC and all the data sources involved in order to maximise the process of implementation of data handling tools.





The scope of data collected should be adapted to the specific nature and research profile of the RDMC.

4.2. Collecting medical data

The new RDMC will need to be prepared and competent to use many distinct types of data sources, differing not only in nature but most of all in the level of digitisation.

Minimum scope of data collected from own sources:

- Hospital information systems (HISs)[3], in particular:
  - demographic data,
  - diseases (ICD10) – underlying and concomitant,
  - procedures performed (ICD9),
  - medications used (active substance, EAN code),
  - hospitalisations and visits (at least date of admission and discharge),
  - event onset times,

[3] Data originating from clinical trials must be so labelled.

  - descriptive data – medical records unrelated to the service (e.g. discharge summaries, doctor's notes, nurse team observations, forms describing patient condition etc.),[4]
- Laboratory diagnostics (lab test results),
- Imaging diagnostics (descriptions and images),
- Pathological diagnostics (descriptions and images),
- Information on the availability of biological material kept in a biobank,
- Genomic analysis (whole genome sequencing, whole exome sequencing).

It is recommended to collect data from other sources:

- Omic (metabolomic, proteomic, transcriptomic) study results,
- Other units/labs cooperating with the RDMC,
- Medical telemetry – records from medical devices and remote monitoring equipment (e.g. holter, activity trackers),
- Individual well-being, lifestyle and/or quality of life patient surveys based on standardised international questionnaires such as SF-36, EQ-5D, HRQoL-14.

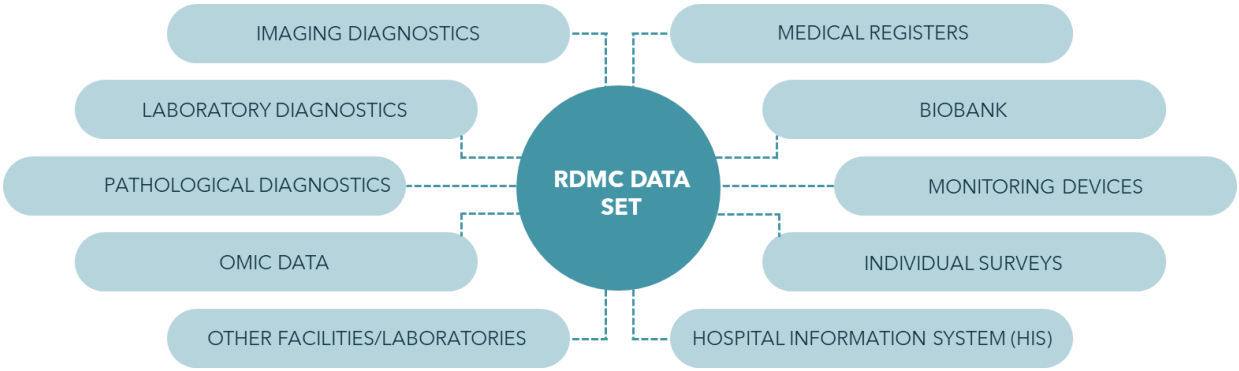
The minimum scope of data for volunteer donors is specified in **Annex no. 1 to the RDMC Standard – Informed Consent Form**.

All the above types of data should be appropriately anonymised or pseudonymised and labelled with the same unique patient code. The scope of both anonymisation and pseudonymisation for the purposes of cooperation between stakeholders (university, teaching hospital, biobank and other) shall be specified by the RDMC. At the RDMC level, patient data should be pseudonymised. Data shared by the RDMC with other RDMCs and the MRA must be anonymised, unless other provisions (e.g. the laws on medical registers) authorise sharing unredacted data. This will ensure consistency of all data stored at the central storage unit.

[4] If the descriptive data are entered in a text field, and not a structured form, they should be entered as a textual value.

5. PRINCIPAL FUNCTIONALITIES

5.1. Regional level (RDMC)



A. An RDMC's IT system must make it possible to:

- process, scale, manage, and search the data being collected, as well as perform statistical analysis thereof,
- generate reports and summaries,
- handle all types of data – structured and unstructured,
- select and index data in order to accelerate data processing,
- use unified dictionaries,
- ensure pseudonymisation and anonymisation depending on the needs of the entity,
- ultimately enable integration with the central system by giving access to an application programming interface (API),
- ultimately exchange/share data with other RDMCs in line with the applicable legislation,
- develop and incorporate new tools and functionalities thanks to the system's modular structure,
- ensure security of the data with the applicable safeguards.

B. An RDMC's IT system is required to:

- use AI algorithms,
- feature software that enables machine learning (ML),
- enable designing mathematical and statistical models.

C. An IT system must be created (a doctor assistant module) that would structure the data introduced to the HIS.

The doctor assistant module will save doctors' time when normalising and structuring online medical records. With its help, the doctor's note will be fully analysed as soon as it is entered. On the basis of the free text typed in by the doctor and recognised by the engine, the assistant should be able to make suggestions on how to make a sentence relevant, clear, unambiguous and comparable regardless of the language/dictionary/terminology it employs or refers to. Ensuring consistency of the medical terminology will enable full exchangeability of the data collected in doctors' medical notes.

The assistant should feature/enable:

- structured voice and text recognition[5],
- integration with IT systems at medical facilities,
- using complex scores such as RECIST 1.1 and PIRADS,
- own templates or adaptable templates from an open access library,
- cloud-based and/or on-site data storage.

D. An RDMC's IT system is required to have a Feasibility module.

- The module must enable consistent presentation of data originating from different sources (adding abstraction layers to imported data) by using unified names of attributes and combining groups of data into defined types of items such as diagnosis, hospitalisation, test, doctor, and it must make it possible to easily find patients using a given query.
- The scope of data imported from different sources must be consistent with the minimum scope of data defined in the standard (the module should ensure pseudonymisation and anonymisation of data depending on the needs). The module must be equipped with a mechanism enabling generation of queries without having specialised training in handling databases, programming or statistics. A user with basic medical knowledge should be able to generate a query.
- The module must be capable of generating criteria for defining and searching patient groups in the database.
- The module must enable searching patient groups on the basis of any defined treatment regiments.

[5] Automatic extraction of EAN, ICD-9, ICD-10 codes and SNOMED CT concepts from unstructured documents.

- The module must make it possible to display all other events that have taken place between two time-dependent events. The module must make it possible to narrow down the search results to a given time period, e.g. last month, last quarter, any date range etc.
- The module must enable defining the search criteria (including treatment regimen parameters).

6. REGIONAL DIGITAL MEDICINE CENTRE STANDARDS (REQUIREMENTS)

6.1. Infrastructure

6.1.1. Equipment and technical resources

As part of project implementation, the entity is to provide the infrastructure necessary to collect, process and secure data. In order to establish and develop an RDMC, the entity is required to propose an individual development concept for the RDMC based on the identified and defined needs in terms of infrastructure and technology. As a result of implementing the RDMC project, the entity should have:

- Servers enabling the effective operation of the RDMC system,*
- Disk space adapted to the RDMC system specifications,*
- Other equipment necessary to ensure effective operation of the RDMC system, defined on the basis of a needs analysis and implementation analysis,
- Premises for the management staff and operators of the RDMC,
- Premises for the IT equipment of the RDMC,
- Other necessary premises to ensure effective operation of the RDMC.

* *Cloud solutions are acceptable*.

6.1.2. RDMC office infrastructure

The objective of the project is to implement this standard at the newly established Regional Digital Medicine Centre. The envisaged number of rooms and the amount and types of equipment purchased must correspond to the entity's needs. The entity must ensure the infrastructure necessary for the RDMC staff to perform their work. The entity must ensure a sufficient number of rooms and work stations for the RDMC staff to perform work and handle the data stored at the RDMC.

6.1.3. Biobank

As part of the RDMC, the entity may establish or equip a biobanking facility, which may be a separate unit independent of the RDMC but working closely with it with a view to improve the quality of the information collected and of the data processing performed at the RDMC. The organisational and functional relationship between the biobank and the RDMC will depend on the internal organisational context and the entity's projects, and shall be at the discretion of the latter. This will make it possible for the information obtained from biological material collected from one patient to be used multiple times in different studies. An

important element of sample deposition is providing sample history, i.e. the conditions in which it has been collected, transported and stored, in accordance with the applicable biobanking standards. It is recommended to collect at least two types of biological material, e.g. blood, biopsy, urine, sputum etc., depending on the type of study and the research profile of the biobank. Peripheral blood/serum is considered the primary type of material. In case of cancer patients there is an additional requirement to biobank samples of cancer tissue (e.g. tumor specimen) in an amount sufficient to enable sequencing, as long as the patient agrees and it is clinically possible to acquire this material. This will allow correlating the changes caused by a given medical condition across different tissues/body fluids and identifying new disease biomarkers.

On the basis of the identified needs, the entity will develop the infrastructure and data analysis methodologies for biological samples originating from clinical research, biobanking facilities and volunteer donors. The established infrastructure and policies must comply with the handbook *Quality Standards for Polish Biobanks, v.2.0* as well as the applicable legislation. Compliance with international standards will also foster international cooperation, which will enhance the research potential and make it possible to publish results in high impact factor journals in the preferred Open Access formula.

In establishing, developing or equipping a biobank, the entity must be guided by the principles described in the handbook *Quality Standards for Polish Biobanks, v.2.0*.

The end result of establishing and/or developing biobanking facilities should be the design and validation of their operations in a way that allows unifying SOPs across different biobanks within one DMC network. The biobank will be required to agree to undergo external certification to confirm conformity with the handbook *Quality Standards for Polish Biobanks, v.2.0*.

6.2. Staffing

6.2.1. RDMC staff

RDMCs will create new job positions depending on the needs and the level of advancement of the RDMC's parent entity.

The RDMC should ensure efficient operation by integrating, analysing, processing and sharing collected data depending on the RDMC's scientific profile understood as the entity's therapeutic areas of research focus. The manager of the RDMC should have the necessary

knowledge and experience in developing new IT tools, including using them in combination with advanced molecular biology methods, supervising and participating in tests or IT system implementations.

The staff should comprise employees with four types of competence:

DATA ENGINEER	DATA CONTROLLER	BIOSTATISTICIAN	DATA SCIENTIST
Person responsible for appropriate collection and processing of raw data, assessing the usability of new sources of information, and designing and setting up new relational databases to store and process incoming information.	Person responsible for how data are stored, processed and shared in relation to the objectives set out by the entity.	Person proficient in using statistical tools to solve problems in the areas of medicine, public health or biology. Experienced in handling analytical algorithms (for primary and secondary analysis) and omic research results.	Expert in analytics responsible for data analysis and interpretation for the purpose of decision making. Combines elements of mathematics, statistics and computer programming. Uses advanced analytical techniques such as machine learning and predictive modelling, as well as scientific approach.

Types of competence do not equal full-time positions. One employee can combine several types of competence or it may be necessary to employ several people with one type of competence; the Applicant should be able to justify the anticipated headcount taking into consideration the time necessary for the RDMC to achieve full operability. The entity determines the number and types of job positions to be created at the RDMC, which should ensure efficient operation of the RDMC, with the minimum of the 4 competences listed in the table above. Employed first should be the analyst, then the architect. The data scientist should only join the team once the architecture is fully operational.

Besides the foregoing, the staff at the RDMC should comprise experts with the following competences, without limitation:

- creating analytical models describing IT systems (logical database models, object and service logic models – API),
- general knowledge of relational databases and SQL, creation and development of online applications, familiarity with scripting languages (e.g. R. Python),
- ability to perform biostatistical and computer analysis,
- ability to develop analytical documentation at the stage of concept definition, IT systems production and implementation design, working with programmers, testers and IT system architects,
- ensuring the operation of IT infrastructure,
- familiarity with the principles of data safety and data processing.

Meeting RDMC's objectives requires a suitable number of qualified staff within an effective and ordered organisational structure. The employment strategy should be based on an ongoing analysis of the RDMC's situation and development plans, and it should take into account the personal needs of individuals in the context of the tasks performed by the RDMC.

6.2.2. Biobank staff

The biobank should develop and implement a human resources management procedure. The organisational structure of the biobank may differ depending on the nature of the entity and the resources it has at its disposal. The biobank should have an approved organisational chart which defines the relationships between individual organisational/functional units. The organisational structure of the biobank should comprise a unit/person responsible for the Quality Assurance and Management System (QAMS).

The biobank should report to a Scientific Committee/Council or a similar advisory body as regards the ongoing research, technical or administrative projects.

The parent entity of the biobank should provide a sufficient number of qualified personnel to ensure smooth running of all the processes performed at the biobank in a continued way after a project is completed.

The competences required of the interdisciplinary team must be defined, as well as scopes of responsibility and working conditions, including workplace risks. The qualifications of an employee should comprise education, completed training or professional experience suitable for the scope of responsibilities applicable at the biobanking facility.

The biobank should define the responsibilities and powers of the employees, specifying the necessary replacements so as to ensure continuity of all the processes at work within the biobank.

In order to guarantee smooth operation of the biobank, it is recommended to employ an interdisciplinary team to ensure efficient operation of the facility and enable certification in accordance with the *Quality Standards for Polish Biobanks, v.2.0*.
The manager of the biobanking facility should have the necessary skills to ensure comprehensive supervision over the facility's operations, including without limitation: initiating, performing and coordinating biobanking-related research activities, defining the facility's research potential, steering its development, shaping the service offer for potential contractors, fostering partnerships in research and development. As regards staffing at the biobank, the entity must be guided by the principles described in the handbook *Quality Standards for Polish Biobanks, v.2.0*.

6.3. Staff training

6.3.1. RDMC staff

Competent staff is a prerequisite for the RDMC's success, which is why training must be planned before the RDMC begins operation. Trained RDMC personnel must understand their scopes of responsibility (see table above) and be able to cooperate with one another in these areas. Training must be hands-on, carried out at the target entities (CTSC, hospital, university), and draw from existing, real-world data instead of simulations. Theoretical

training should be treated as bare minimum. Procedures should also be in place to prevent high turnover of staff, given that training will be time-consuming.

It is recommended to develop and implement a policy regarding induction and periodic training, as well as other activities aimed at improving the skills of RDMC staff. Training may focus on:

- introducing a new method to the process,
- introducing a modification to an existing technological process,
- consolidation (reminder) of knowledge about the technological processes used,
- acquiring new knowledge in areas related to or associated with biobanking,
- quality and risk management.

The RDMC should provide each employee with on-the-job training regarding the applicable SOPs and other procedures, as well as give them opportunities to follow additional professional training programmes. Participation in third-party training programmes should always be confirmed with a relevant document, e.g. a certificate, diploma or attendance certificate, as provided by the organisers. For internal training, an attendance list is sufficient.

6.3.2. Biobank staff

The biobank should develop and implement a periodic training policy. The facility should provide training to every employee (internal training, i.e. provided by another employee, biobank manager or his/her designees, or external training provided by independent third parties with the necessary expertise). The biobank should develop and keep updated a valid training plan. The training plan should be prepared by a person responsible for quality management and approved by the manager of the biobanking facility. The approval of the training plan must be confirmed e.g. with a signature and date of approval.

The biobank should periodically test the knowledge of its employees who perform the technological process to make sure that in doing so, they follow the internal procedures.

Participation in third-party training programmes should always be confirmed with a relevant document, e.g. a certificate, diploma or attendance certificate, as provided by the organisers. For internal training, an attendance list is sufficient. Copies of documents confirming employee participation in training should be stored by the manager of the biobanking facility or by another person of authority. After internal employee training, the

biobank should test the knowledge of the employees using tools such as a written test, oral questions or a practical skills test.

6.3.3. Learning curve

The RDMC's parent entity should suggest ways to verify the staff's skills associated with the relevant area of competence of the CTSC (thematic scope, types of data) by developing a model of checking and documenting competence acquisition. Importantly, when planning further development of the CTSC and RDMC, it will be necessary to plan full training and induction of each new hire, which means that a buffer time will need to be envisaged for the new employees to adapt before they can take up their responsibilities. Therefore, the entity must allow for some delay resulting from the learning curve and present ways of mitigating the risks arising from this.

6.4. Business continuity

In order to ensure business continuity, the RDMC should make provisions to ensure recovery of operations after a disaster or other disruptive incident (e.g. a cyberattack). With a view to ensuring continuity of operations, the entity is required to develop a Business Continuity Plan including Disaster Recovery Plans which include:

A. Providing for restoration from a secure backup copy.

B. Meeting the assumed level of RDMC service after restoration.

The RDMC should ensure:

- maintenance and safe operation of the IT system,
- physical and environmental security, including access control,
- security and continuity of services fundamental to the operation of the RDMC,
- backups to enable post-failure recovery:
  - In order to maintain business continuity, e.g. in the event of a failure, backups should be made of all systems that constitute the service provided, in accordance with the pre-established schedule and priorities.
  - These backups should be tested periodically by being restored in a dedicated recovery environment; they should each time undergo mandatory verification of correctness and include recovery reports.

- To ensure security, backups should be created regularly regardless of whether the environment is on-premise or cloud-based (e.g. on permanent tapes or other back-up systems). For on-premise environments, the copies should be delocalised in case of a fire in one of the processing centres. It is recommended to apply the 3-2-1 rule for backups. The rule is to keep at least three copies of data (3 x back-up), with two copies stored on different devices (2 x back-up), one of which (1 x back-up) is located away from the company's headquarters, e.g. in the cloud. Back-up copies should be encrypted.

6.5. Data security

One of the principal ideas of the Call is the safety of patient medical data. In securing and processing research data, including clinical study data, RDMCs will have to operate in accordance with the legal basis developed by the European Union. This legislation calls for special protection of the data, since the context of their processing may pose a serious risk to fundamental rights and freedoms. The RDMC system must ensure strict control of access to the systems where the data are stored and processed, as well as maximum data protection with the selected solution. In the context of cybersecurity, it is permissible to store data on the servers of suppliers or service providers that meet the regulatory requirements, whichever technical solution they use (cloud or traditional server).

After production start-up or joining the DMC network, the RDMC system must meet the requirements of OWASP ASVS at least at the L1 level, which should be confirmed with a report from penetration tests carried out by an independent company specializing in application security. It is recommended that such penetration tests be performed at least once a year.

In addition, the RDMC system should meet the following requirements:

- The data must be protected against unauthorized access by means of user rights regulation.
- Each system user (employee of the university, teaching hospital, biobanking facility or other entity involved in the development of the RDMC) must have a separate login and password.
- Accessing any system functionality (regardless of the number of modules) will only be possible after logging in (with double authentication).

- The system should only be accessed from office devices authorized by the administrator.
- It must be possible to add, block, delete and edit user accounts.
- Compliance must be ensured with the Regulation of the Council of Ministers of 12 April 2012 on the National Interoperability Framework, minimum requirements for public registers and electronic exchange of information, as well as minimum requirements for ICT systems.
- For critical data entered manually, additional checks are required via validation tools according to risk-based analysis.
- Anonymisation and pseudonymisation of the data collected must be possible, depending on the needs of the entity.
- An appropriate level of cybersecurity must be applied, ensuring data flow security in accordance with the applicable law, current recommendations of Centrum e-Zdrowia for building cybersecurity systems, and the recommendations of the European Union Agency for Cybersecurity (ENISA) for the medical sector.
- The system must comply with the legislation applicable as at the launch date and be promptly adaptable to any changes in this regard.
- Compliance must be ensured with the National Cybersecurity Standards (NSC).
- The system must log out or lock user sessions after a specified time of inactivity.
- The system must comply with the Regulation (EU) No. 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR) and it must be developed in accordance with the principles of Privacy by Design and Privacy by Default. Data minimisation must be guaranteed. The system must allow individuals to exercise their rights, e.g. the right to rectify and access data, the right to data portability, and it must provide the obligatory information under Articles 13 or 14 of the GDPR.
- The accountability of the users' activities in the system must be guaranteed and all user activity tracked – every data entry/modification and the original data along with the information on who entered them and when. Once entered, no data may be permanently overwritten and therefore irreversibly lost.
- The system must create and keep a system log for at least 6 months, recording all system users and their activities, with an option to analyse the history of changed data values.

- The administrator must have access to the Event Log, which will include a log of all logins, changes, revocations and deletions/erasures of data.
- The system must be protected against unauthorized access. The security measures must operate at the levels of client (application) and server (database server).
- The system must enable data export or migration to a new server environment if it is necessary to modernise or replace equipment.

An important aspect of secure data sharing is access to the data by project contributors. This access should be adequately controlled:

- Users and individuals should only have access to the data made available to them by the controller of that data.
- Every data set stored on RDMC servers must have a designated employee responsible for the data.
- Data can be transferred between units, provided that a relevant approval of the Ethics Committee has been obtained.
- System users must be trained in the security policies in force at the RDMC and undertake not to store any special categories of data on their private computers.

Therefore, the structure of the system must allow for secure sharing of data with other entities, particularly public entities. In the case of commercial entities, after submitting a relevant request and obtaining approval (granted by the project manager and the relevant Ethics Committee), it will be possible to grant access to sensitive data (only within the RDMC system).

6.6. Minimum output data quality to enable aggregated data analysis (file types, quality, formats and sizes)

The entity must ensure high quality of data in the data warehouse, both at the level of data definition and architecture, as well as content. The quality of data definition should be understood as the degree to which data description corresponds to the actual concept or object, so that the data can be understood the same way without ambiguities by whoever has access to them. This may refer to the data description complying with generally applicable standards or norms. The quality of data architecture defines the quality of the data model and databases in a way that enables shared access by all authorised recipients.

High-quality data architecture ensures:

- presence of appropriate relationships across the data (data consistency),

- stability of data models from the point of view of access by various data recipients, regardless of the access applications used,
- flexibility of data models in terms of modification or expansion.

The quality of the data content is determined by the degree of comparability and accuracy of how the collected data reflect reality and meet the expectations of the recipients. Measures of data content quality include, without limitation:

- consistency of the data content with the definition (do the data describe the reality they are intended to describe, e.g. the data do not meet this criterion if they are ambiguous),
- completeness of the data (are there any missing values),
- compliance of the data content with the applicable rules (correctness of the data, e.g. whether they are consistent with the available model data),
- accuracy of data content,
- no duplicate data.

In order to integrate the above-mentioned data, they must first be collected in one safe place prepared for analysis (within one RDMC). It is necessary to ensure such management of project data (scientific research, development works, experiments) that will allow for the transfer of data between projects, provided a relevant patient consent and Ethics Committee approval are obtained. A suggested, but not required solution of this problem is to store data in separate virtual machines or disk instances, which, if necessary (subject to approval of the Ethics Committee), can be installed with other projects and connected with each other.

6.7. Introduction of a standard for medical data

The standards of data collection must not limit the exchange of data. The standards selected by the MRA are the FAIR principles (Findability, Accessibility, Interoperability, and Reuse).[6] The RDMC should store medical events and EDMs created, at least to the extent specified in the Regulation of the Minister of Health of 26 June 2020 on the detailed scope of data regarding a medical event processed in the information system and on the manner and dates of transferring these data to the Medical Information System, as amended.

The RDMC should use exactly the same indexation of medical event resources as the source database P1 in accordance with the standards established in P1. The extension of the

[6] https://www.go-fair.org/fair-principles (accessed on 12 January 2023)

information versus the scope specified in the law should be standardised by the RDMC as part of cooperation between different RDMCs. Other data attributes for patient events should be stored according to the logic compliant with the HL7 CDA standard. One standard should apply to all centres. This will ensure the interoperability of RDMC systems, and subsequently create a potential for interoperability with the entire e-health system in Poland and Europe. The target standardisation of data will be carried out within the DMHO (MRA).

5.7.2 Imaging data

Imaging data should be saved as DICOM files and stored on servers or in the cloud, connected to the hospital infrastructure that handles imaging examinations: diagnostics, radiology, histopathology and other. The RDMC should have systems for image anonymisation and pseudonymisation, depending on the specific needs and legal requirements.

The data must include the following information:

- the acquired image as a DICOM file (a requirement of the e-Health Standards),[7]
- reference to the unique identifier of the patient tested,
- the type of test performed,
- the standard (national or European) used to perform the examination,
- information whether the data are raw or processed (if processed, then how),
- the diagnosis.

6.7.1. Sequential data

Sequential data saved in textual formats (compressed or not) often contain genetic variants, which are patient-specific and sensitive. These are high-volume data and require efficient computing servers at the stages of quality control and processing. The Applicant will need to demonstrate its capability to both secure and process the data.

The data must include the following information:

- the acquired raw sequences as fastq files,
- reference to the unique identifier of the patient tested,
- the type of test performed,
- details of the sequencer used to perform the test,

[7] https://www.gov.pl/web/ia/standardy-e-zdrowia (accessed on 10 February 2023)

- contact information of the person responsible for the test.

In addition, the processed data should be able to be stored, provided that the protocol used to create them is adequately described and can be reused.
A VCF is a text file format (saved as compressed file). It contains meta-information lines, a header line, and then data lines, each containing information about the position in the genome.
The quality of *vcf (Variant Call Format) files should be defined in accordance with the current VCF/BCF specification. The group leading the management and expansion of the format is the file format team of the Global Alliance for Genomics and Health Data working group. The major version of the specification (and its updates) can be found at https://github.com/samtools/hts-specs
The latest version of the specification is v4.3. The main measure of the quality of a sequence in a vcf file is the QUAL value.
The International Genome Sample Resource at https://www.internationalgenome.org/ can be used as a reference data resource.

6.7.2. Other data

Other laboratory, metabolomic, proteomic, and other omic data are usually in textual table formats but in some cases they can use other formats specific to medical equipment manufacturers.

These data must include the following information:

- the type of data obtained,
- the standard (national or European) used to perform the test,
- reference to the unique patient identifier.

6.8. Data sharing and exchange

Due to the increasing number of risks related to special categories of data, it was decided to implement a federated solution with a potential central solution to be implemented in the future. Such a solution consists in storing data at the entity that has obtained the data and has the legal authorisation to store and process them, in accordance with the GDPR and national laws on special categories of data.

An additional element of the federated system is the introduction of relations and standards of data exchange between data collecting units and the central unit. For Poland, the central unit is Centrum e-Zdrowie. Another candidate for the central unit will be the Digital

Medicine Head Office (DMHO) at the MRA, once it is operational and legally established. Within the federal structure, data collectors will inform the central unit if they have specific types of data (for example, clinical study, imaging, or sequential data), but will not forward all of this data to the central unit.

It should be emphasized that technically, transferring data is possible, but it is done only in justified cases and after obtaining the appropriate consents at the legal and bioethical level. It is possible to create a communication network based on an appropriately secured and encrypted SSH-2 protocol, which will enable connecting and transferring the data. The Agency recommends using the Fast Healthcare Interoperability Resources (FHIR) standard. FHIR can be used to transfer data collected in medical systems to the extent covered by the standard.

6.9. Ethical and legal aspects

The RDMC should operate in accordance with the principles of ethics as well as national and international laws applicable in Poland with regard to handling human biological material and data, with due respect of the dignity and fundamental rights of research subjects and ensuring unbiased organisation, management and risk analysis.

The RDMC should document its activities by archiving documentation such as permits, decisions, opinions and notifications. It is also recommended to specify the duration of storage of documentation and related data. The relationship between the RDMC and its parent entity must be defined. Information on the adopted rules of conduct with regard to the ethical and legal aspects important from the subjects' point of view, for example on the manner of transferring samples and data to other entities, is posted on the websites of the entities involved in the development of the RDMC, with reference to specific documents. Information intended for research subjects should be provided in a simple language understandable to the average reader.

In accordance with Article 35 of the GDPR, where the processing of personal data is likely to result in a high risk to the rights or freedoms of natural persons, a Data Protection Impact Assessment (DPIA) should be carried out before the processing is initiated to identify the threats that may affect personal data and apply appropriate preventive measures and security mechanisms that preclude or minimise the risk of data breaches.

7. ANTICIPATED RESULTS AND BENEFITS

1. Creation of an Application Programming Interface (API) to further develop the system; integration with the central system by making the API available.
2. System openness. The system must be expandable with the addition of new modules.
3. The system must enable data export or migration to a new server environment if it is necessary to modernize or replace any equipment.
4. The biobank created, developed or equipped holds certificates in accordance with the standard of the handbook *Quality Standards for Polish Biobanks, v.2.0*.
5. Support of study design (e.g. power calculations, implementation of digital tools) – improved quality at the design stage (quality by design),
6. Creating a clinical and omic database on whose basis digital tools can be generated in the form of prognostic, predictive and therapeutic artificial intelligence algorithms.
7. Real-time data analysis – improved patient safety and quality of results.
8. Post-research data analysis – data quality controlled from baseline to end of study and reliable results with a high publication potential.
9. Storing data from clinical trials conducted under contracts with the MRA and supervision of their integrity and security in cooperation with the MRA.
10. Data processing based on international standards and establishing good practices in cooperation with the MRA.
11. Unifying the standards of analysis and the quality of the data collected.
12. Standardisation of data from different HISs.
13. Standardisation of data from different biobanking facilities.
14. Establishing solutions for the anonymisation/pseudonymisation, integration, security and transfer of data in compliance with the applicable standards and legislation in cooperation with the MRA.
15. Targeted transferring of data within the Network, including to the Digital Medicine Head Office at the MRA, for secondary analysis.
16. Developing standards for the disclosure of data to third parties with a view to improving data availability in close cooperation with and under the leadership of the MRA. The standard will be implemented after the legislative changes.
17. Developing the infrastructure and data analysis methodologies for biological samples originating from clinical research, biobanking facilities and volunteer donors.
18. Promoting the digital medicine programme abroad and establishing international partnerships focused on standardisation and data analysis.

Annexes:

1. Informed Consent Form