DATA MANAGEMENT PLAN

1. DATA SUMMARY

The ***SALVAGE***-DMP describes the data management life cycle for the data to be collected, processed, and generated during the project period, and beyond. It further outlines the methodologies and describes standards applied, as well as how the data will be shared, made openly accessible after curation, and ensuring preservation during and beyond the life cycle of the project. Within *SALVAGE* project the majority of data will be generated *de novo* by 7 project beneficiaries, i.e. in 7 different Czech Organizations (partners) and some multiomics data will be re-used from available data repositories. All data to be generated in the *SALVAGE* project, partners consulted and agreed in the DMP. During the implementation of the project the DMP will be updated.

Re-using existing data

The *SALVAGE* project will reuse data as follows: 1) Genomic and somatic DNA sequences for modeling and targeting design, 2) phenotyping data from characterized genes, 3) RD- literature in PUBMED and 4) various transcriptomics, 5) exposomics and other omics data from available repositories, including the rare disease registries, Czech genome/multiome project database, the National Cancer Registry and the National Health Information Registry. Using and reviewing these data resources are regular part of the project work.

Types and formats of data generated and their reusability

Different types and formats of data will be generated in the project. Standardised data formats will be used as much as possible. Formats of raw and processed files for the different data types are listed in Table ***1***.

Table 1: Data types and formats

Type of data	Partner where data is generated	Raw data format	Processed data format(s)
Genomics incl. metagenomics	IMTM-UP, MMCI, UHM, UHO, RECETOX, UO	.fastq.gz	.bam, .vcf, .tsv
Proteomics	IMTM-UP, MMCI	.raw	mzTab, .csv (peptide/protein lists from Proteome Discoverer)
Transcriptomics	IMTM-UP, UHM, UO	.fastq.gz	.bam; preprocessed data formats: .html, .txt, .zip, .csv, .sf, .tsv, .pdf, .r and .gtf
Exposomics	RECETOX	.mzML	.raw/.lcd, mzTAB - standardized ISA-Tab file
MethSeq	IMTM-UP	.fastq	.bam; .bedGraph
Metabolomics	IMTM-UP, UPc, RECETOX	.raw	.mzML, After pre-processing: .tsv files (MAF)
Cellular, tissue and organismal phenotyping	MMCI, IMTM-UP, UHM, UO	.xml	.html; .json
Images	MMCI, UHO, UHM, UO; IMTM-UP	.tiff .dicom	.html
Patient-DBs & registries	MMCI, UHO, UHM, IMTM-UP, RECETOX, UO, CU	.xml .dicom tiff	.html; .json, DASTA, HL7-FHRI (communication formats)
Alphafold – protein modelling	IMTM-UP, RECETOX, MMCI, UO	.fasta	.pdb

Majority of data will be reused by the project partners as well as they can be reused by broad research community irrespective whether academic or industrial. All the data from model generation, and their characterization will be searchable and accessible under relevant web-portals (e.g. https://portal.imtm.cz, www.genasis.cz).

The purpose of the data generation and re-use and its relation to the objectives of the project.

The purpose of the data generation is to establish validated models of (pre)cancers suitable for designing and testing early cancer therapies and diagnostics and monitoring disease development and therapeutic response. Once the models will be established a and validated, data will be generated based on characterization of the models, biomarkers description, design and effect of the proposed therapy. It would be o values of the research community when the model and data will be re-used and further expanded.

Expected size of the data to be generated or re-use.

The expected size of the data of different types are listed in

Table 2. The total expected size of raw data based on previous knowledge is estimated to be ~300 TB, including the images (CT, MRI, PET/CT, bioluminescence/fluorescence, flow cytometry, etc.). The size of pre-processed data will be approximately the same. There will be additional data for integrated omics analyses with major data volume coming from whole genome sequencing analyses. Nonetheless, processed data will be small in comparison to the raw data.

Table 2: Expected size of the data generated within the project

Type of data	Expected size of raw data (per sample if not stated otherwise)	Expected size of processed data (per sample if not stated otherwise)
Proteomics	10 GB	tbd
DNA-seq.& genomics	2 GB – 2 TB	>5 GB
transcriptomics	>5 GB	15 GB
Exposomics	Non-target: per file ~100MB, mzML ~200MB target: ~200KB	~200-1000 KB
Metabolomics	~1 GB per sample ~130 GB per analytical batch (~135 samples)	~800 MB total size for .tsv files
phenotyping	45 GB (cohort of RD model) 20MB (10 000 data files)	500 MB (cohort of RD model) 20MB (10 000 data files)
Images	1 GB (cohort of RD model)	100 MB (cohort of RD model)

Data origin/provenance (either generated or re-used)

Considering genome sequencing data, they are coming from an existing research project from partners. The anonymization and/or pseudonymization procedure is only performed by the medical partner involved, all other project partners/analyzing sites receive the information already in pseudonymized format. Any patient data transferred to the analysis sites is done in a pseudonymized format associated with a unique ID only. Regarding the patients/study participants data, they are visible as a system assigned Patient ID only, which does not relate to any personal information provided. System assigned Patient ID is unique across all studies and registries in the ClinData system, proprietary developed at the IMTM-UP. There are no anonymisation techniques used, as in case of discovery of clinically relevant results, those are reported to the patient/study participant. Similarly, data on activity of candidate drugs/therapies tested for *in vitro* activity will be managed using portal MedChemBio and *in vivo* data in PreClinData software solution. The population cohort data of CELSPAC are available in RECETOX RI.

I. New data. Majority of data will be generated - see the *Table 1 & 2* . These data will be derived from animal and cellular models, from which the cells will be mostly of the human origin. Environmental data will be derived. Human subjects from a cohort of healthy or diseased individuals recruited in studies; participants or their legal representatives have provided the required informed consent covering the future use for research purposes.
II. previously generated data - reused. These data consider the Patient/study participant data/ case report forms, that are accessible only through ClinData software and/or hospital information systems. Primary personal data, coming from patient/study participant sample analysis are associated with a unique ID only (pseudonymised). In ClinData software, personal information is visible only to the authorised users. Users can be authorised in accordance with GDPR practices.

Data utilisation outside of the project

The generated data sets (including the metadata) together with the generated models of disease will be useful to broad biomedical academical community as well as pharma and biotech companies. The data linked to the validated (pre)cancer models will serve users and clients who intend to study cancer development, compared various models, or treatment modalities as well as to compare the model-patient situation. Data from biomarker and/or innovative therapeutic studies may be also used for medicinal product approval process by competent (regulatory) authorities and implemented to clinical practice. Newly collected data combined with the established databases can be used also for AI-approaches to explore disease development and treatment design with its effectiveness.

2. FAIR DATA / FAIR DATA

MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA / ZAJIŠTĚNÍ DOHLEDATELNOSTI DAT (FINABILITY), VČETNĚ USTANOVENÍ PRO METADATA

Human and machine-readable metadata will be added to allow data sets to be ***findable***. Dedicated research data repositories hosted by European Bioinformatics Institute (EBI) provide stable and unique identifiers for each submitted data set. If possible, data sets will be assigned a Digital Object Identifier (DOI), which will make them easily citable. One possibility for automatically assigning a DOI is submission to Zenodo. Due to legal reasons, some data cannot be uploaded to EBI repositories at the moment. We anticipate federated data storage solutions such as Federated EGA or EOSC-CZ. Therefore, data will remain at the secured ClinData, PreClinData or MedChemBio repositories, and made findable through sharing metadata via federated tools (e.g. FAIR Data Points or Beacons.)

Naming conventions and ontologies

Standardized terms (ontologies) will be used as much as possible. Existing metadata schemas such as the Dublin Core Standard will be used. Standardized naming conventions are currently developed e.g., for genomics data in the FAIR genomes project (ZonMw 846003201), the Global Alliance for Genomics & Health (GA4GH,), the Beyond 1 Million Genomes, the HUPO Proteomics Standards Initiative (PSI).
A codebook with definition of a metadata schema has been published. Similar efforts for standardizing metadata for other types of omics data are being pursued by the Netherlands X-omics Initiative (NWO 184.034.019). We will build upon these developments to report the data generated in *SALVAGE*. Keywords will be provided to allow optimal re-use possibilities. These keywords will be selected from controlled vocabularies and deposited alongside the metadata. Data sets will be provided with a version number and/or time stamps.

Data identification & identifiers.

Standard vocabularies and identifiers will be used for all data types wherever possible. See *Table 3* for a list of used vocabularies and identifiers. The usage of such vocabularies and identifiers has been already established for the work with animals models. Standardization efforts in cancer research are currently being solved within the EOSC4 Cancer consortia, where Dr. Hajduch (IMTM-UP) leads the WP2 focused on data harmonization and ontology. The plan is to follow or modify nomenclature and ontology already developed at the NCI NIH. Within the *SALVAGE* project we will follow procedures standardized within the EOSC4 Cancer, which should be available in late 2023.

Provision of rich metadata to allow discovery

As the *SALVAGE* project uses multiple technologies and basic research project, the *SALVAGE* platform will focus to manage and make available rich metadata obtained from processes that are highly standardized and reproducible. As mentioned above, this involves the whole pipeline for cell and mouse model generation and comprehensive phenotyping, and the preclinical testing, high-throughput screening, omics approaches and preclinical testing performed by project participants. Since there are three medical research infrastructures involved in the project (EATRIS, BBMRI, ECRIN), they will take care about the future data sustainability in the project.

Searchable metadata: optimized for discovery, harvesting and re-using

In the area of the multiomic studies, disease registries, cellular and animal model generation and their standardized phenotyping as well as the preclinical testing of therapies the IMTM and RECETOX will be serving as a technological and data hub for all other partners and interacting third parties, the data and metadata will be harvested, stored, and processed in way allowing searching and re-using the data. For this purpose, the RECETOX ***and IMTM databases and software dealing data and its interactive presentation*** will be further expanded to make the metadata visible. The basic categories presented are as follows:
1) ***Multiomic and Phenotype data Quality-controlled***, aggregated, analyzed and processed will be displayed to anyone with an internet connection, for free; 2) ***Bulk Data***, which could be used by collaborators aiming to use them in their own custom queries, can be accessed by FTP, API and Batch Query for all types analysis needed; 3) ***Advanced analytical Tools*** are provided via EATRIS-CZ consortium – those tools help view, use and analyse our data; 4) ***Data Collections*** - we shall adapt our pages to provide curated pages that highlight certain areas of our data and make them searchable and interoperable.
A codebook with definition of a metadata schema has been published for types of omics data (the Netherlands X-omics Initiative. We shall build upon these developments to report the new data. Keywords will be provided to allow optimal re-use possibilities. Keywords will be selected from controlled vocabularies and deposited alongside the metadata. The **metadata will be offered in such a way that it can be harvested and indexed.**

MAKING DATA ACCESSIBLE / ZPŘÍSTUPNĚNÍ DAT (ACCESSIBILITY)

REPOSITORY / REPOZITÁŘE (ÚLOŽIŠTĚ)

Data deposition and repositories

The ***data flow*** in the *SALVAGE* project contains two major IT/informatics systems allocated in the 2 LRI, the Masaryk University (RECETOX) and IMTM – these institutions represent national nodes for three sustainable research infrastructures (BBMRI, EATRIS, ECRIN, EIRENE) have already built bioinformatics units, which not only have repositories for all the data they generate but the data are FAIR and interoperable international. The IMTM informatics will be now adapted to the *SALVAGE* project in the way that all data generated will be sustainable beyond the *SALVAGE* project.

Data repositories and data identification

IMTM has already built own repositories for all the work needed for the chemical biology, animal disease models, collection and mining of clinical and multiomic data and currently develops solutions for data visualization and analysis. To increase data reproducibility, the IMTM team is currently developing also laboratory information and management system, all located at https://portal.imtm.cz and "omics" data under the umbrella of EATRIS-CZ; thus, these data are also interoperable.

Identification of the digital object

The digital objects equipped with an Digital Object Identifier (DOI) are harvested in the standardized biomarker and/or phenotyping pipelines and stored in the IMTM repository or delivered to other internationally recognized repositories, which can be accessed using various data processing tools. Moreover, GDPR sensitive data will be uploaded to the Digital Research Environment (DRE) will be transferred by the Azure Storage Explorer or similar solutions at the IMTM. The DRE complies with GDPR. Additionally, any user accessing the *SALVAGE* DRE workspace and its data is required to sign a terms and conditions of use agreement ensuring data protection under GDPR.

DATA

Data – openly available and restricted access

Principally, data generated from the standardised and validated models and procedures will be presents in an Open-access form although an **embargo** will be established to allow researched publish the new finding. These data will be provided in an open access, after their curation, at *SALVAGE* partner institutions or by informatics departments of the international consortia (EATRIS, BBMRI, ECRIN). Data that will create a part of ***intellectual property (IP)*** will be handled according to the rules of Transfer offices of a particular Organisation (Partners) of the project and could be released only if the would not endanger the IP and its protection. The embargo for data release will be generally 24 months after the data curation and will require agreement by the Steering committee of the *SALVAGE* project.

Free and standardized access protocol

Key data generated within the standardised process of cell, tissue pathophysiology and animal model data, (pre)clinical studies will be accessible through a standardized access protocol upon registration of a user, unless specific rules/restrictions apply for IP or GDPR sensitive data.

Restrictions for data usage

Access to restricted data will be provided through a ***Data Access Committee (DAC)*** and by a ***Data Access Agreement (DAA)*** between a data user and the DAC. Repositories such as the European Genome-Phenome-Archive facilitate this process. The DAC of *SALVAGE project* includes all members of its *Steering committee*. The generated molecular, diagnostics, and (pre)clinical data is considered sensitive personal data that is potentially identifiable and will require a Data Access Committee in accordance with the GDPR. A machine-readable license will be provided. The identity of the person accessing the data will be ascertained through the signatures of institute's representatives on the DAA.

Evaluation of the access to the data & data access committee

In general, the access to the project's omics and phenotypic data will be defined in ***SALVAGE access policy*** on data access requests – it will be created and presented on our LRIs websites till December 2024. In addition, controlled access to sensitive personal data will be evaluated by the Data Access Committee. The ***identity of the person accessing the data*** will be ascertained vie registration form accessible on the IMTM websites.

METADATA

Collecting metadata is established for all processes regarding animal model generation, archiving, and comprehensive phenotyping or (pre)clinical development, which is also supported by ISO 15189, ISO 17025 and good laboratory practice and good clinical practice certified workplaces of the *SALVAGE* project partners and by involved large research infrastructures (EATRIS, BBMRI, ECRIN). The IMTM LRI (member of EATRIS, ELIXIR, BBMRI, BioImaging and OpenScreen) will take care about establishing the metadata standard used in these European infrastructures. Existing metadata standards will be re-used as much as possible in all standardize experimental procedures, i.e. a model generation, its characterization, preclinical testing of therapeutic approaches and biomarker identification and validity.

Availability of metadata

The metadata obtained will be made openly available and licensed under a public domain dedication CC-0 after publication besides it creates an IP or GDPR issues, which will be settled up according to the particular partner institutional rules and in line with law. Typically, the metadata will be provided after the registration of an applicant.

Data/metadata availability and findability: Long-term preservation and curation

Concerning all animal and cell models, all metadata will be guaranteed to remain available together with all other data generated. Regarding the human/patient data, due to legal concerns about uploading sensitive personal data to servers outside of the Czech Republic or even outside of the European Union, data cannot be submitted to access-controlled dedicated repositories from EMBL-EBI (the European Genome-phenome Archive (EGA), PRIDE Archive, MetaboLights). Ongoing developments on federated data storage solutions might solve this issue in the future. Therefore, all data including raw and processed omics data will remain at the secure IMTM ClinData repository. Potentially identifiable information will have access controlled by an Executive Committee.

Data accessibility – specific software

Raw data formats for some data types require proprietary software to access. Where possible, we will use open formats that do not require proprietary software. Documentation and/or links about/to relevant software will be included. Data analysis tools and pipelines will be made available via publishing and appropriate licensing. Patient/study participant data/ case report forms, are accessible only through ClinData software.

MAKING DATA INTEROPERABLE / INTEROPERABILITA DAT

Data and metadata vocabularies, standards, formats and methodologies for data interoperability, exchange and re-use within and across disciplines.

The generated data will be made available using standardized formats commonly used in the respective domain (see Table 1). Existing metadata standards, such as the Dublin Core Standard, will be used to make the data interoperable. If relevant, mouse model generation and phenotyping will follow the IMPC- EMPReSS (see in the Table 3) standard with embedded standardised statistics and metadata. Omics-specific metadata standards include standards of the Metabolomics Standards Initiative (MSI), Minimum Information About a Next-generation Sequencing Experiment (MINSEQE) guidelines, Minimum Information About a Proteomics Experiment (MIAPE), Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE), as well as guidelines from commonly used domain-specific data archives (EBI MetaboLights, EBI European Nucleotide Archive – ENA). Standard vocabularies will be used for all data types where possible. See *Table 3* for a list of used vocabularies and identifiers.

Ontologies and vocabularies used

Standard vocabularies will be used for all data types where possible. See *Table 3* for a list of used vocabularies and identifiers. In case uncommon or project specific ontologies or vocabularies are used we will provide mappings to more commonly used ontologies.

Table 3: Controlled vocabularies and ontologies used in SALVAGE project platform. For all gene-related identifiers, the human genome reference build GRCh38/hg38 is used.

Data type	Identifiers, ontologies, controlled vocabularies
Metabolites	HMDB ID, ChEBI ID (Chemical Entities of Biological Interest)

Measurements & units identification	Units of Measurement Ontology (UO)
Phenotypes	Human Phenotype Ontology (HPO), National Cancer Institute Thesaurus (NCIT) EMPReSS: https://www.mousephenotype.org/impress/index
Genes, transcripts	HUGO Gene Nomenclature (HGNC), Ensembl Gene ID, Ensembl Transcript ID
Gene annotation	Gene Ontology (GO)
Genomic coordinates	Unique identifiers based on chromosome (referred to via GenBank ID and version) and genomic coordinate (GRCh38) corresponding to, e.g., CpG site
Peptides and proteins	UniProtKB Sequence and UniProtKB Accession Number
microRNAs (sequencing and qRT-PCR)	miRbase ID (release 20)
Sample materials	Schema developed by FAIR genomes project
Analysis information	Schema developed by FAIR genomes project
Protocols, methods, experimental metadata	Ontology for Biomedical Investigations (OBI), Chemical Methods Ontology (CHMO), Experimental Factor Ontology (EFO), NCI Thesaurus OBO Edition, Metabolomics Standards Initiative Ontology (MSIO), PRIDE Controlled Vocabulary
Roles and contributions	CRO - Contributor Role Ontology

Data and qualified references to other data and resources

The data generated and analysed will use qualified references – such a referencing has been established in the framework of international consortia BBMRI, ECRIN, EATRIS , EIRENEor are being established in specific disease situations, for instance EOSC4Cancer, etc.

INCREASE DATA RE-USE / ZVYŠENÍ OPAKOVANÉHO POUŽITÍ DAT

System & documentation needed to validate data analysis and facilitate data re-use

To effectively manage complex translational medicine data, IMTM/EATRIS-CZ developed several proprietary tools for data stewardship for in vitro, preclinical and clinical data. The tools are available to broad research community and users on daily basis:

Administration module: Provides authorization and authentication of all users on via IMTM/EATRIS-CZ data portal. It is central authentication server with single sign on and two phase authentications support.

ClinData: Software solution designed for data management of clinical trials, clinical registries, various healthcare or scientific databases. The ClinData is currently used for daily management of single/multicentric clinical trials and registries (close to 50.000 patients in more than 50 clinical trials, including the Czech genome/multiome projects).

PreClinData: Software solution designed for stewardship of preclinical animal data. PreClinData include also simple biostatistcs module to evaluate safety and efficacy of experimental therapies on daily basis (survival, clinical signs, tumor volume, body weights, histopathology, etc.). PreClinData hosts above 150 animal studies of multiple users.

MedChemBio Portal: LIMS for medicinal chemistry, high-throughput screening and chemical biology. It includes compound registration and management, QA, in vitro biology, pharmacology, data analysis, storage, export and reporting. The portal is primarily used for analysis of in vitro biological activity of small molecules for collaborating chemical groups. There are registered >130k biologically active small molecules tested for biomedical applications.

CovIT: Cloud-based laboratory management and information system CovIT was developed in response to COVID-19 pandemic for laboratories involved in diagnostic PCR testing. The system includes full capabilities of LIMS with automatic reporting of cases to the National Registry of Infection Diseases and self-reporting of epidemiologically relevant contacts .

Above mentioned systems and tools will be further adapted and expanded for the specific purpose of *SALVAGE* project.

Data availability in the public domain for widest re-using

Relevant data generated within the standardized process of cellular animal model generation, phenotyping will be available via IMTM web-portals after publication to allow their free and wide re-using. The preclinical data may be a subject of IP issues/rules and thus, such data will be subjected an internal control of the Data Access and steering Committees in compliance with an institutional Transfer office, regarding their release. Clinical data will be available via IMTM portal solutions with access limited by the GDPR and/or IP rules.

Data usage by third parties

Majority *SALVAGE* data generated in standardized pipelines for cellular and animal model generation will be available after publication/termination of the project to third parties, unless restricted by IP and/or GDPR rules.

Data quality assurance processes.
Quality Assurance of the data is inherent to all *SALVAGE* data generated in standardized pipelines for cellular and animal model generation, their comprehensive phenotyping and preclinical studied including the omics analysis. Quality Control measures will be available together with the data and metadata. Integrity of data files will be secured by storing an md5 checksum (or related measure) with the data.

Fair principles & Data security
Especial focus on data security is dedicated to personal and patient data. Personally identifiable information will be stored ClinData system developed by the IMTM. It is specifically designed to provide a secure environment storing and accessing this kind of data. The data are protected by two-phase authentication and authorization. The data are backed up daily to facility located on the premises of UP Olomouc under surveillance and controlled access. The transferred data are fully encrypted. The system is compliant with all legislation required to manage clinical data, which includes extensive regulation of personal data management. Personally identifiable information such as genomics data will require a ***Data Access Committee***.

3. OTHER RESEARCH OUTPUTS / DALŠÍ VÝSTUPY VÝZKUMU

Management of project outputs
The feasibility plan of the *SALVAGE* project contains a plan of all research outputs – i.e. not only publications but also patents, and other forms of IP protection. Upon successful completion of result protection, the results and data will be released to support the usage either the IP and/or serve the users, and of course to bring treatment benefits to patients. The governance of the project, with respect to the partners involved (and their internal rules) will make freely available all the protocols, data resources as well as primary and secondary material.
In addition, the *SALVAGE* will supervise the reaching the outputs planned in The Feasibility plan of the project and moreover, in tight interaction with the four LRIs involved, RECETOX-MUNI, MMCI and IMTM, will strive to convert the new knowledge, models and technologies into services provided by these large research infrastructures.

4. ALLOCATION OF RESOURCES / ALOKACE ZDROJŮ

Allocation of resources and capacity for making data and other research outputs FAIR
The *SALVAGE* project is built as a consortium of specialized laboratories of major Czech Universities and university hospitals and three medical Large Research Infrastructures (BBMRI, ECRIN and EATRIS) and environmental RI (RECETOX RI), where the MMCI, MUNI, and IMTM-UP function also as national nodes. Thus, dealing with big data set for last decade, these three LRIs, have already established procedures, capacity, technologies, and management that deal with FAIR data. Particularly EATRIS and BBMRI developed and implemented data management tools and processes for comprehensive preclinical and clinical studies. IMTM, RECETOX and MMCI also maintain a continuous and indispensable effort in integrative bioinformatics as part of its involvement in phenotyping research by large-scale analysis of multiomic datasets and image analysis. The Bioinformatics Units covers handling of all phenotyping data including quality control, statistical analysis, and their storage and uploading them into public web interface, where can be accessed by global community. The Patient Data management will be primarily operated by several medical partners that has established agenda and operation system with dealing with patient and their data (MMCI, UHM, UP, UHO).

Budget for data management and FAIR data processing
Due to the involvement of 4 LRIs, most of the budget for data management incl. all the process will be cover by these three LRIs, which also includes operation of the computing and data storage servers. Moreover, we also plan to participate in future calls of EOSC-CZ (OP JAC) focused on development of national data repositories.
SALVAGE institutions have already built their bioinformatics/IT department as well as capacity, software for the data management and analysis, and thus *SALVAGE* project reserved specific budget for bioinformatics work dedicated only to this project such upgrading the omics databases, disease model databases with its analytics, and pathogenic protein variant interactions, etc. Regarding the patient data, the budget for this management is within the medical partners and covered from the routine management of hospital/laboratory management information systems.

Responsibility for data management in the SALVAGE project

In the *SALVAGE* project, there is ***Executive Committee***, in which data stewards of and heads of each Research intent participate; the DMC is led by the ***Scientific manager***. The institutional data stewards are responsible for the whole data management agenda of the respective institute/LRI. The directors of the LRIs are responsible for securing sufficient budget and capacity for *SALVAGE* project.

Long-term data preservation

MMCI/BBMRI, MUNI/RECETOX, MUNI/ECRIN and IMTM/EATRIS-CZ, will take care about long-term and sustainable data storage and managements they have long-term development plans including the investments into the IT, and they also get supported from the EOSC-CZ projects.

Trusted repositories for long term preservation and curation & Data security and protection provisions

SALVEGE involved LRIs has built own data management systems with repositories that are also interlinked with international repositories. In the case of EATRIS, it is collecting data from all the working modules/portals, and after the configuration they are processed in two parallel branches. First branch is dedicated to store and display data on our local server and web. The second branch is dedicated to data transfer from collaborating academic and collaborators hospitals to the IMTM data portals. To cope with the data management, the IMTM has invested into own servers as a primary site for initial data storage and analysis – these servers and repositories aims in long-term operation.

5. ETHICS / ETIKA

Ethics or legal issues that can have an impact on data sharing

Regarding the ethical aspects of the project, there are ethical and legal aspects of data sharing:
Aspect No. 1: The informed consent forms, clinical study registration and study protocol approved by the local ethics committees.
Aspect No. 2: Processing of personal data that include research data, considered sensitive data. Within the framework appropriate forms were collected from each project partner (including third parties) to determine their scope of personal data processing for the project, as well as ascertain which institutional policies are in place.

Informed consent for data sharing and long-term preservation

The *SALVAGE* management, i.e. ***Data management committee (DMC)*** will adapt the informatics and IT systems in the way that informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data.

6. OTHER ISSUES / OSTATNÍ

The DMP will be updated accordingly after receiving the funding and reviewed yearly to adapt it based on the project and output development as well in a reaction to development regulation.