RESEARCH DATA MANAGEMENT PLAN

ARIS 2.0 FORM

Document Version: 1.1
last update: 31.08.2026

0	General information	
0.1	**Project code**	J5-70192
0.2	**Project name**	Scientific Basis for the Use of Virtual Reality in the Modulation of Biopsychosocial Factors That Limit Exercise Therapy (VIRTUES)
0.3	**Project start and end dates**	1. 3. 2026-28. 2. 2029
0.4	**Description of the project from the point of view of handling research data**	The VIRTUES project examines the scientific basis of the use of immersive virtual reality (VR) to modulate biopsychosocial factors that limit the implementation and participation in exercise therapy in people with osteoarthritis of the knee. The theoretical starting point is a biopsychosocial model of pain and functional disability, according to which participation in exercise is influenced by pain, kinesiophobia, self-efficacy, body image, beliefs about the disease and user experience, in addition to physical impairments. The project connects UP FVZ and the Jožef Stefan Institute and includes the development of adaptive VR solutions, pilot studies of acute effects and user experience, and a randomized controlled study to evaluate a 12-week VR-based exercise and educational intervention. Quantitative, qualitative and technical data will be collected: demographic and clinical data, questionnaires, pain and kinesiophobia assessments, kinematics, EMG, pressure center forces and movements, muscle strength, functional tests, physical activity data, intervention logs, user experience data, interviews and focus groups, and data on the performance of VR software and hardware. The data will be used for solution development and optimization, feasibility checks, statistical and thematic analysis, comparison of conditions or groups, and preparation of scientific publications. During implementation, data will be pseudonymized, and direct identifiers will be kept separately. Metadata, documentation, protocols, analysis code, and only those anonymized or aggregated data where the risk of re-identification is acceptable will be openly published. Sensitive health, biometric and qualitative data will be restricted in accordance with ethical approval, consents and legislation.
0.5	**Project manager code**	21495
0.6	**Name and surname of the project manager**	Prof. Dr. Nejc Šarabon

0.7	**Name and surname of the person in charge of support in handling research data in the RO**	Assoc. Prof. Ana Slavec, Ph.D. (UP Open Data Expert Support, NRRP and FAIR)
0.8	**Name and surname of the person responsible for handling research data in the project**	Prof. Dr. Nejc Šarabon
0.9	**E-mail address of the person responsible for handling research data in the project**	nejc.sarabon@fvz.upr.si
0.10	**NRRP version and date**	Version 1.0, 27/08/2026
1	**Summary and description of research data**	
1.1	**Will you reuse existing data (your own or someone else's) in the project?**	☒ Yes ☐ No The project will use data already collected and the results of related previous research in people with osteoarthritis of the knee who participated in a structured exercise program. These include demographic and clinical characteristics, data on pain, physical activity, degree of radiological disability and kinesiophobia, as well as qualitative interviews, transcripts and derived codes or thematic categories. The findings are used for the selection of exercises, the design of avatars and virtual environments, and the preparation of educational content. Reuse will be limited to the extent allowed by the original informed consent, ethical approval and legal basis; new informed consent will be obtained for new research procedures. Existing scientific literature and publicly available technical documentation of VR hardware and software will also be used.
1.2	**What types of data will you create or reuse and in what formats will it be stored?**	Record and administrative data: Lists of potential and enrolled participants, signed consents, screening forms, randomization schemes, visit logs, intervention logs, adverse event records, and a link key between identity and research tag. Electronic records will be in DOCX, PDF/A, XLSX and, if necessary, CSV; signed originals will also be kept in paper form. Quantitative personal, health and behavioural data: Demographic, anthropometric and clinical data, diagnosis and degree of KOA, pain, kinesiophobia, self-efficacy, knowledge and beliefs, body image, user experience, technology adoption, VR side effects, quality of life, physical activity, exercise participation, and results of functional and movement tests. The working formats will

		be XLSX, CSV and SAV; mainly CSV/TSV with a separate data dictionary in TXT/PDF/A will be prepared for archiving. Biomechanical and physiological data: Raw and processed data from 3D kinematics, EMG, inertial and depth sensors or cameras, force measurement plates, dynamometry, electronic goniometry and physical activity meters. The raw data will be preserved in the original device formats; the processed data and key time series will be exported to open or widely used formats (presumably C3D, CSV/TSV, TXT and MAT) along with information on sampling frequency, filtering, units and processing procedures. VR Development Specifications: Camera configurations, lighting, background and clothing, tracking quality scores, system logs, coordinate and skeletal data, avatar and virtual environment parameters, configuration files and software versions. The formats envisaged are XLSX/CSV, JSON, TXT/LOG, Unity/C# project files and, if necessary, PNG/JPEG and MP4 or other native video format for RGB-, depth or thermal imagery. Qualitative data: Audio recordings of semi-structured interviews and focus groups, verbatim transcripts, field notes, coding schemes, codebooks and printouts from ATLAS.ti. Working formats will be WAV/M4A, DOCX/TXT and the source format of ATLAS.ti; for longer-term usability, anonymized exports to TXT/CSV/XML/PDF/A will be selected. Audio recordings and full transcripts will not be made public. Documentation, analyses and results: Research protocols, SOPs, data dictionaries, README, processing log, analysis scripts, tables, images, reports, publications and presentations. DOCX, PDF/A, TXT/Markdown, CSV, SPSS syntax and R, Python or MATLAB scripts if necessary; images and graphs will be in PNG/SVG/PDF. Archival copies will be in open and independent formats whenever possible.
1.3	**What is the expected size of the data you plan to create or reuse?**	☐ 0-10 GB ☐ 10-100 GB ☒ 100-1000 GB ☐ >1000 GB
2	**Storing and backing up data**	
2.1	**Where will the data be stored and backed up during the execution of the project?**	The primary working copy of the research data will be stored in a dedicated project folder on the institutionally managed UP/FVZ network or server storage; the data generated by the JSI's partner will be stored in a comparably secured institutional environment of the JSI

		until the agreed secure transfer. Access will be limited to named members of the research group using individual accounts, multi-factor authentication where available, and the principle of least necessary privileges. Operational editing and administration will be carried out by Manca Opara Zupančič together with the leaders of individual researches, and supervision will be carried out by the person responsible for data handling and the project manager. Direct identifiers and the link key will be stored separately from the research data, in a separate encrypted folder with a narrower circle of access. Signed consents and other paper documents will be stored in a locked cabinet on the premises of the UP FVZ. On measuring computers and devices, raw data will be stored only temporarily; after successful encrypted transmission, integrity verification and backup, it will be removed. Transfers between UP and JSI will be carried out via approved encrypted institutional channels, not via ordinary e-mail or personal cloud accounts. Transfer media will be used only exceptionally and must be encrypted.
2.2	**In which trusted repository will the published data be in accordance with the principle of "open as much as possible, closed as necessary"?**	The primary institutional repository will be the Repository of the University of Primorska (RUP; https://repozitorij.upr.si/), which will publish metadata, documentation, protocols, other research results, and anonymized or aggregated data that meet the conditions for open publication. Databases related to individual scientific publications will be published, where appropriate, in an appropriate trusted repository, such as Zenoda. For individual health, biometric, audio and full qualitative data, as a general rule, at least a public metadata record describing the restriction will be published, while the files themselves will be closed or accessed through a controlled process. If the selected repository does not allow such access, permanent administration will be provided on an institutionally secure repository until the expiry of the permitted retention period, and the metadata, data dictionary and access conditions will be published in the repository.
3.	**Provision of data in the FAIR manner through repositories**	
3.1	**Will all data be open access, respecting the principle of "open as much as possible and closed as necessary"? If**	☐ Yes ☒ No The metadata of all data publications and, where ethical approval, consents, copyright and risk of re-identification, anonymised or aggregated data tables, data dictionaries,

	not, please explain the specifics of access.	codebooks, protocols, analysis scripts, technical documentation and non-sensitive derived data will be openly published. The data directly supporting the scientific publication will be published or recorded in the repository at the latest at the time of publication of the publication and cited in the article with a permanent identifier. Open access will be restricted to direct identifiers, link keys, medical and biometric data, individual time series, raw RGB, depth or video recordings, audio recordings, full transcripts, and small combinations of data that are at risk of re-identification. The restrictions are based on the protection of personal data, ethical approval, the content of informed consent, the confidentiality of qualitative statements, and any intellectual property rights and third-party licenses. For data with restricted access, the metadata record and the declaration of availability will be publicly available. Justified requests for access will be handled by the project manager and the Data Officer until the expiry of the retention period, in consultation with the Data Protection Officer and the Ethics Committee, if necessary. The applicant will have to define the purpose, the legal and ethical basis, the required scope and security measures, and sign an agreement on the use of the data. Only the minimum necessary anonymised or pseudonymised set will be transmitted through a secure channel. After the expiry of the retention period, access to the destroyed individual data will no longer be possible. A time-lock will only be applied when it is necessary for the review process, the protection of intellectual property or the preparation of the primary publication, and will be as short as possible.
3.2	**How will you enable the usability of your research data (metadata, keywords, software, licenses, others)?**	Each database will be accompanied by a README file with the research title, authors and responsibilities, project code and funder, version and date, purpose, design, pattern, time and environment of collection, equipment and program used, processing steps, inclusion and exclusion criteria, missing data, known limitations, and instructions for use. A file manifest, checksums, a data dictionary with variable names, labels, codes, allowable values and units of measurement, as well as links to the protocol, publication, and analysis code will be added. For qualitative collections, a description of the context, a guide to the interview or focus groups, anonymization rules, and a code book will be prepared. File and folder names will be systematic and without personal names, such as VIRTUES_WP3_STUDY01_P001_T1_EMG_RAW_v01_2027-03-15.ext. Dates will be written according to ISO 8601, units

		of measurement will be written according to SI, text exports to UTF-8, and missing values will be uniformly marked and explained. Raw data will be immutable, and processed data and derivatives will be separated and linked to the processing log and script version. Metadata will be entered in accordance with the scheme of the selected repository, usually with DataCite or Dublin Core elements: title, authors with ORCID, description, keywords, funder and project code, temporal and spatial coverage, methods, type of source, related publications, version, license and access conditions. Content keywords such as virtual reality, osteoarthritis of the knee, exercise therapy, kinesiophobia, pain, kinematics, electromyography, user experience and rehabilitation will be used. For non-sensitive data and documentation, a CC BY 4.0 license or other appropriate open license is envisaged, for metadata a repository license, and for software code an open source license, which will be determined after the ownership arrangement and verification of dependency licenses. For data with limited access, instead of an open license, an agreement on the use of the data and clear access conditions will be indicated. Proprietary formats will be accompanied by an open export, a description of the software, a version and, where possible, a conversion script.
3.3	**What research data quality assurance procedures will you use?**	The collection will be carried out according to pre-prepared protocols and SOPs. Prior to the start, researchers will be trained in recruitment, informed consent, the use of questionnaires, the performance of functional and biomechanical measurements, VR procedures, the recording of adverse events, the naming of files and data protection. The measuring equipment will be checked, calibrated and synchronized before use; the measurement log will record the device used, serial or system designation, setting, software version, contractor and any deviations. Technical validation of VR tracking will systematically check factors such as angle of incidence, background and clothing color, illumination, distance and height and tilt of cameras, number and placement of cameras, and type of movement. Standardized tests and predefined tracking error estimates will be used; decisions on selected settings will be documented with reasons and configuration version. Electronic forms will use predefined data types, required fields, allowable ranges, and logical controls. Duplicate identifiers, discrepancies between visits, extreme or impossible values, units, dates, randomization, completeness, and match to source records will be checked after entry. Critical variables and manual entries will be

		independently validated, and corrections will be recorded without overwriting the original values. Raw data will be kept unchanged and read-only; cleanup, filtering, and calculations will be performed on copies with reproducible scripts and a change log. Analysis code, VR configurations, and documentation will be version-driven. Reasons for measurement exclusions, artifacts, missing data, protocol deviations, and participant withdrawals will be documented. Audio recordings will be transcribed against the transcript, identification information will be removed, and the encoding will be reviewed by at least one other person; the evolving coding structure will be discussed in the research team. Prior to archiving, a final review will be carried out for completeness, anonymization, documentation, file formats, licenses, and consistency between data, analyses, tables, and published results.
4.	**Ethical and legal aspects**	
4.1	**Will you process, store, create or reuse sensitive personal data or special types of personal data during the implementation of the project?**	☒ Yes ☐ No The project processes direct identifiers and contact data, demographic and anthropometric data, data on the diagnosis of KOA, pain, chronic diseases, medications, physical activity, psychosocial factors, function and quality of life, and biometric data on movement, muscle activity and other bodily responses. Audio recordings, full transcripts and possible RGB, depth or video recordings can allow direct or indirect identification. The research involves patients and healthy adults. The processing will be carried out in accordance with the General Data Protection Regulation, ZVOP-2, internal acts of UP and JSI, the approved protocol of the Commission of the Republic of Slovenia for Medical Ethics (No. 0120-160/2026-2711-3) and other applicable ethical approvals for reused data. Prior to inclusion, written informed consent will be obtained, and participants will be informed of the purpose, scope of processing, recipients, retention period, rights, possibility of withdrawal and the intended manner of publication of the results. Only the minimum amount of data necessary for research objectives will be collected. During the survey, the data will be pseudonymised with research tags. The link key will be stored separately and accessible only to a limited range of people. Prior to sharing or publication, the data will be further anonymised or aggregated; for qualitative data, personal, geographical,

		professional and other indirect identifiers will be removed or generalised, and verbatim quotes will be reviewed for identifiability. Full anonymisation will only be used when the link to the individual can no longer be effectively restored by reasonable means. Identification and research data will be kept for a maximum period of time permitted by ethical approval and informed consent; the current documentation provides for a maximum of five years after the completion of the research in question, followed by permanent deletion or destruction. Direct identifiers and audio or video recordings will be deleted sooner than they are no longer needed for quality checks, data withdrawals or analysis. In the long term, only data and results that are no longer personal and mandatory project and archival documentation with deadlines under the applicable institutional rules may remain.
4.2	**How will data security and the protection of sensitive data be taken care of during the survey?**	The main risks are unauthorized access, loss or theft of the device, misidentification of identity with research data, disclosure in a small sample, voice or image recognition, interception in transit, accidental upload to the provider's cloud services, modification or corruption of files, and use of data outside of the approved purpose. Risks will be mitigated through pseudonymisation, separation of identification and research data, encryption in transit and on devices, individual accounts, multi-factor authentication, restriction of privileges, regular system updates, backups, access logs and documented deletion. VR glasses, cameras and accompanying software will be set up in such a way that, where possible, data is processed locally and is not automatically synchronised to consumer or provider clouds. Before using third-party services, the contractual legal basis, the location of processing, subcontractors and data transmission will be verified. Data will not be sent by ordinary email or stored in personal accounts. Personal, medical or pseudonymised data will not be entered into publicly available generative AI tools or other unauthorised online tools; only data where the identification has been irreversibly removed and the use complies with internal rules is allowed. New members will sign a confidentiality statement and undergo training prior to access. Suspicions of a data breach will be immediately reported to the project manager, the ICT department responsible and the Data Protection Officer and dealt with through an institutional procedure; where appropriate, a notification will also be made to the supervisory authority and participants. A documented re-identification risk assessment will be carried out prior to publication or transmission.

5.	**Other research results**	
5.1	**Will you create or reuse other research results in addition to the data?**	☒ Yes ☐ No The project will create VR software and digital components: Unity/C# software code, algorithms for real-time capture and processing of kinematic data, feedback mechanisms, camera and sensor configurations, virtual environments, avatars, audio-visual and interactive elements, technical tests, validation records, and installation and user instructions. The code and configurations will be managed in a controlled version repository with marked releases, change log, dependency documentation, and backup. Research protocols and SOPs, questionnaires and interview guides, educational content on KOA and pain, exercise protocols, codebooks, data dictionaries, analysis scripts, randomization schemes, work package reports, publications, conference presentations, professional papers, and a doctoral dissertation will also be produced. Physical biological samples will not be collected; physical results are primarily limited to signed consents and other mandatory documentation. Non-sensitive protocols, documentation, educational content, analysis code and those parts of the software for which the consortium has the right to dispose of will be published in the RUP or an appropriate repository and provided with a version, a permanent identifier and a license. Components containing personal data, security-sensitive information, trade secrets or third-party elements will be excluded or replaced with a description, schema or open case without protected content.
6.	**Financial resources**	
6.1	**What will be the costs of handling data and other results of the project according to the FAIR principles and how will they be covered?**	Direct costs include researchers' time for planning and documenting collections, preparing forms and protocols, data entry and verification, transcription, pseudonymization and anonymization, preparation of data dictionaries and metadata, reproducible analyses, quality checking, preparation of repository packages, and handling access requests. Also included are the potential costs of additional secure server capacity, encrypted media, archiving of extensive sensor or video data, SPSS, ATLAS.ti, Unity and other specialized software licenses, and technical support. Indirect costs include the use of the institutional ICT infrastructure of the UP and the JSI, central backup, information security, support of a data protection officer, a data consultant, a library and a repository, and regular

		maintenance of measuring equipment. Most of these costs will be covered by the existing infrastructure, overhead costs and working hours envisaged in the project work packages, especially in the management, development, data processing and statistical analyses. Any additional storage or curation services will be financed from the material costs of the project or other institutional resources after prior approval by the project manager. The need for additional resources will be reviewed at the time of each major database and before the start of a long-term intervention, when the volume of raw sensor and video data can be more accurately estimated. The costs of data handling will be monitored as part of regular project reporting.

Document Version: 1.1

last update: 31.08.2026