Form for Submitting Research Data to RUNG

TITLE OF THE RESEARCH DATA: _____

WERE THE RESEARCH DATA CREATED AT ANY SLOVENIAN PUBLIC RESEARCH ORGANISATION?
☐ yes
☐ no
☐ not specified

ARE RESEARCH DATA PUBLISHED AND EQUIPPED WITH PID (PID is a unique persistent identifier of a digital object, e.g., DOI, Handle ...)?
☐ yes
☐ no
☐ not specified

LINK TO THE RESEARCH DATA:
https://_____

FORMAT OF THE RESEARCH DATA
☐ completed dataset[1]
☐ corpus[2]
☐ dataset/data series[3]

TYPE OF RESEARCH DATA
☐ aggregated/summary data[4]
☐ clinical trial data[5]
☐ compiled and collected data[6]
☐ coded data[7]
☐ experimental data[8]

---

[1] A research data collection is the result of research and is usually published as a dataset in a national or international trusted data archive (repository). It must be publicly accessible and documented in such a way that it enables the replication of published scientific findings. Its quality is assessed based on the data, metadata, and comprehensive accompanying documentation (e.g., measurement instruments, research process reports, sampling information, consent forms, protocols, software; the content may also be described in a ReadMe file when the repository lacks detailed metadata fields). Its scientific significance is reflected in its usefulness for researching a broad range of applied or theoretically grounded problems.

[2] A corpus is a special type of research data collection: a unified, standardly annotated, and internally structured set of authentic texts created according to predefined criteria and for a specific purpose, available in electronic form and equipped with tools that enable multi-layered search and statistical processing of data.

[3] A dataset represents partial, incomplete research data published in a data archive (repository), thereby fulfilling FAIR principles and enabling research reproducibility. The data are accompanied by metadata, a license, a persistent identifier, and annotations that clearly indicate to users that the dataset is not complete.

[4] Statistical data referring to broader classes, groups, or categories. The data are averages, sums, or other derivations from individual-level data, such that it is no longer possible to distinguish characteristics of individuals within those classes, groups, or categories. Examples: the number of unemployed people and their age in certain geographic regions, or national-level statistics on the occurrence of certain offences originally derived from statistical data from individual police districts.

[5] Data from a research study in which an individual or several people are prospectively assigned to one or more interventions (which may include a placebo or other control), on the basis of which the effects of these interventions on biomedical or behavioural health-related outcomes are assessed.

[6] Data collected or compiled from multiple, often heterogeneous sources that share one or more common reference points, where at least one of the sources was originally prepared for another purpose. These data are integrated into a new entity. Example: providing data on the number of universities over the past 150 years using various available sources (financial documents, official statistics, university registers), combining survey data with official statistical data on geographical areas (population density, number of physicians per capita, etc.), or using an RSS protocol to collect blog posts or tweets.

[7] Qualitative data (text, video, audio, or image) originally created for other purposes, converted into quantitative data (expressed in matrices by units) using coding techniques in accordance with predefined categorisation schemes. Example: coding political party manifestos, such as in the study of the 2009 European Parliament elections (doi:10.4232/1.10204).

[8] Data resulting from an experimental research method, in which some or all variables included in the hypotheses are manipulated.

☐ genomic data[9]
☐ geospatial data (GIS data)[10]
☐ laboratory notebook, research journal, research notes[11]
☐ measurement and testing data[12]
☐ observational data[13]
☐ recorded data[14]
☐ simulation data[15]
☐ survey/research data[16]
☐ other

If you did not provide this information within RUNG, please complete the following:

PROJECT INFORMATION:

Project name (title): _____

Project code (number): _____

Research organization where the project took place: _____

Project acronym: _____

Funding agencies: _____

Project start date: _____

Project end date: _____

---

[9] Genomic data refer to the genome and DNA of an individual organism. They are used in bioinformatics, where genomes of living organisms are collected, stored, and processed. Genomic data mostly originate from sequencing techniques, but the term is broader than sequence data. They may also include non-sequence-based data, such as microarray data, real-time PCR panel data, and data from pharmacogenomic studies.

[10] Geospatial data are all types of data with spatial coordinates that allow mapping onto the Earth's surface. They may represent physical objects, discrete areas, or continuous surfaces. Discrete geospatial data are usually represented using vector data composed of points, lines, and polygons, whereas continuous geospatial data are typically represented with raster data as a grid of cells, each with its own value. Many applications across various fields generate geospatial data (such as GIS, remote-sensing equipment, GPS units, archaeological and surveying instruments, manual mapping, and computer-aided design (CAD)) in many formats, including images, vectors, text, and tabular data. Vector geospatial data include tables listing archaeological sites with their coordinates, text files (e.g., XML) with coordinates and topology of historical road networks, and voting results for political parties by administrative areas. Raster-based geospatial data include satellite imagery, aerial photographs, scanned maps, and digital maps of elevation/terrain, vegetation, land use, sea-surface temperature, air pollution, soil types, etc.

[11] A laboratory notebook is the primary record of research. Researchers use laboratory notebooks to document their hypotheses, experiments, and the initial analyses or interpretations of those experiments. The term applies to both traditional and electronic laboratory notebooks.

[12] Data resulting from the assessment of specific properties (or characteristics) of organisms, objects, phenomena, and/or processes using established standards and/or specialised instruments or techniques.

[13] Data obtained through observational research, which includes collecting observations in real time (e.g., observing behaviour, events, the development of a (health) condition or disease, etc.) without attempting to manipulate any of the independent variables.

[14] Data recorded by mechanical or electronic devices in a form that allows the stored information to be retrieved and/or reproduced.
Example: images or sounds on a disk or magnetic tape.

[15] Data derived from modelling or simulating real-world processes, events, or systems, often using computer programs.
Example: a program that models how household consumption changes in response to indirect tax changes; or a dataset of hypothetical patients and their exposure to a drug, initial health status, and known adverse outcomes.

[16] Data obtained from a survey in which statistical methodology is systematically applied to a selected sample of a population to collect data and investigate or assess a particular characteristic of that population. This includes population censuses, sample surveys, databases from administrative records and derived statistical activities, and questionnaires.