

Development of erosion prediction tool for sustainable soil management

Vinča Institute of Nuclear Sciences, University of Belgrade
Faculty of Sciences and Mathematics, University of Niš
Faculty of Agriculture, University of Belgrade
<https://www.predict-er.com/>



Deliverable D4.3

Strategy for Knowledge Management and Protection

This research was supported by the Science Fund of the Republic of Serbia, #GRANT No 7047,
Development of erosion prediction tool for sustainable soil management - Predict-Er.



Science Fund
of the Republic of Serbia

Supported by the Science Fund of the Republic of Serbia, Program PRISMA, #GRANT No 7047

PROJECT TITLE	Development of erosion prediction tool for sustainable soil management
ACRONYM	<i>Predict-Er</i>
GRANT NUMBER	7047

DOCUMENT TITLE	Strategy for Knowledge Management and Protection
DELIVERABLE NUMBER	D4.3
RELATED WORK PACKAGE	WP4 - Project management and dissemination
WP LEAD ORGANIZATION	VINS
RELATED SUB-ACTIVITY	4.3 Development of Strategy for Knowledge Management and Protection
VERSION	V0.1
DELIVERY DATE	31/05/2024
FILENAME	Deliverable_4.3_V0.2.docx
DISSEMINATION LEVEL	Public
STATEMENT OF ORIGINALITY	This deliverable contains original unpublished work except where indicated otherwise. Acknowledgment of previously published material and the work of others has been made through appropriate citation, quotation, or both

Contents

1 Introduction.....	4
2 Intellectual Property and Knowledge Management in the <i>Predict-Er</i> Project: Strategies for Dissemination and Protection	4
3 About <i>Predict-Er</i> Data Management Policy	6
3.1 Description of the Dataset.....	6
3.2 Dataset reference and name	7
3.3 Standards and Metadata	7
3.4 Data sharing and access	8
3.5 Data archiving and preservation	9
4 Type of data generated in <i>Predict-Er</i>	9
5 Making <i>Predict-Er</i> data findable	10
6 Ethical and Legal Considerations in the <i>Predict-Er</i> Project: Ensuring Data Integrity and Confidentiality	11

1 Introduction

The knowledge management and protection strategy aims to widely disseminate the knowledge created while protecting the intellectual property rights of project results that can be exploited. The comprehensive knowledge management and protection strategy (D4.3) is consistent with the knowledge dissemination and utilization plan (D4.2, month 4).

Knowledge management and protection ensure the appropriate protection of intellectual property (IP) in the background and foreground and its exchange in an open knowledge environment. Interested parties are granted royalty-free access rights to existing and new project knowledge as required for their work in the context of project activities.

The purpose of the knowledge management and protection strategy is to provide relevant information about the data collected and used by the partners of the *Predict-Er* project.

The *Predict-Er* project employs a multidisciplinary strategy, merging nuclear, analytical, statistical, and remote sensing methodologies to produce detailed in-field data at high resolution. This data will be fused into an advanced Web-GIS tool with the capacity to predict alterations in soil erosion rates and sediment dynamics under various management scenarios. To achieve this, a platform for harvesting and sharing data has been created which is located on the intranet of *Predict-Er* web page (<https://Predict-Er.com/>).

The aim is to provide a strategy for managing data generated and collected during the project and optimize access to and re-use of research data. The strategy is intended to be a living document that will outline how the *Predict-Er* research data will be handled during and after the project. So it will be reviewed and updated when necessary.

The strategy outlines the data management lifecycle for all datasets involved in the project. It includes: i) management of data during and after the project, ii) formats and categories of data collected and generated, iii) approaches and standards to be followed, iv) timing and methods for data sharing or open access, and finally v) data curation and preservation.

Predict-Er will produce a range of outcomes, encompassing measurement data, observations, validation protocols, and scientific articles concerning the efficacy of a collection of innovative solutions aimed at enhancing sustainable land management. These efforts aim to boost productivity while mitigating sediment production.

2 Intellectual Property and Knowledge Management in the *Predict-Er* Project: Strategies for Dissemination and Protection

This section delineates the management of internal knowledge within the *Predict-Er* project.

Internal knowledge management will be facilitated through a web-based secure professional platform for the sharing of information and documents. Each partner will have the possibility to upload key documents and papers.

The primary goal of Intellectual Property Rights (IPR) protection is to safeguard the rights of the partners involved in the project. This protection ensures that each partner has the freedom to utilize their unique strengths and advantages and to fully capitalize on their market positions. When there is a viable opportunity to patent the intellectual property or innovations arising from the project, the project team will pursue patent protection. Additionally, partners will deposit all obtained research data, using both classical and digital means, to provide open access to public data while restricting access to private data.

In the event of conflicts arising regarding IPR within the *Predict-Er* project, partners will adhere to a structured conflict resolution process. Initially, partners involved in the dispute will engage in open and constructive dialogue to identify the root causes of the disagreement and explore potential solutions. This dialogue will be facilitated by designated representatives from each partner organization, ensuring that all perspectives are considered and respected. If a resolution cannot be reached through direct negotiation, the consortium will appoint a neutral mediator or expert to facilitate discussions and assist in finding a mutually acceptable solution. Mediation sessions will be conducted in a fair and impartial manner, with a focus on achieving a resolution that upholds the interests and rights of all parties involved. If necessary, partners may seek legal advice to clarify any legal aspects or implications of the dispute. Throughout the conflict resolution process, transparency, confidentiality, and mutual respect will be maintained to foster trust and collaboration among the partners. The ultimate aim of the conflict resolution process is to safeguard the integrity of the project and its outcomes while preserving the collaborative spirit and positive working relationships among the consortium members.

The knowledge management strategy can be illustrated as follows. To address Open Access (OA), publications resulting from project implementation will be categorized as "green OA." This term refers to a form of OA where authors self-archive their work in online repositories. Researchers will attain Open Access (OA) by self-archiving their published articles or final peer-reviewed manuscripts in online repositories, both before and after the publication process, and in accordance with the license rights and permissions of the respective journals.

Parties agree to cooperate in submitting, examining, publishing, and defending dissertations or theses incorporating project foreground or background. Access rights to project results are granted to concerned partners under reasonable conditions for non-commercial research and teaching purposes.

The consortium adheres to general principles regarding Pre-Existing Know-How (PEKH) or "Background," whereby each partner retains ownership of its PEKH and related intellectual and industrial property rights. Regarding foreground ownership, participants generating the foreground retain ownership, with joint ownership established for the jointly generated foreground.

The management of Intellectual Property (IP) encompasses IP development, protection of background and foreground IP, and exchange within an open-knowledge environment. An inventory of background IP is conducted to establish a foundation for trustful cooperation.

The Consortium aims to disseminate generated knowledge widely while safeguarding exploitable project results' IPR. Dissemination and exploitation activities are planned to achieve both objectives effectively. Background knowledge remains the intellectual property of the providing partner, while basic knowledge is regulated with partners entitled to utilize project results royalty-free to support dissemination and exploitation activities.

3 About *Predict-Er* Data Management Policy

The data management policy will detail the current status of the data being produced. For each set of data collected, processed, and generated in the project the following elements are addressed: (1) Description of the Dataset, (2) Dataset reference and name, (3) Standards and metadata, (4) Data sharing and access, (5) Data archiving and preservation.

3.1 Description of the Dataset

This section details the data to be collected or generated, encompassing its source (in cases of collection), characteristics, scope, potential utility, and contribution to scientific publications, if applicable. It also includes insights into the presence of comparable data, along with opportunities for integration and repurposing. Datasets that will be collected/generated during Project realization are listed in Table 1.

Table 1. Datasets to be gathered or produced throughout the project's execution.

Dataset name	Description	Data used concerning the research field
Slope and river flow direction maps	Shows terrain slope and direction of water flow	For preparing field campaigns and comparison with future data derived from high-resolution DEM
Land cover map from Sentinel 2 satellite imagery	Land cover map derived from Sentinel 2 satellite imagery in 10 m resolution	For comparison with future land cover data derived from high-resolution multispectral imagery
Vectorized geological and soil maps	Vectorized special-purpose maps	Data will be used as a parameter for estimating soil erosion and assessing the soil's vulnerability to erosion
LiDAR high-resolution point cloud	LiDAR produces detailed digital 3D relief model	For obtaining high-resolution DEM
High-resolution DEM	Digital Elevation Model characterized by mm-scale and cm-scale resolution	For obtaining an mm-scale and cm-scale resolution 3D terrain model for feeding erosion process models
Aerial (drone) imagery – RGB and multispectral	Aerial photos that provide a high level of detail (mm-scale and cm-scale resolution and absence of clouds)	For high-resolution orthophoto, spectral indices with multispectral data to obtain land cover maps
Highly detailed land cover	Land cover derived from our multispectral drone imagery characterized by mm-scale and cm-scale resolution	For obtaining mm-scale and cm-scale resolution land cover for feeding erosion process models
Sediment sources and suspended sediments	Physicochemical properties, stable element contents, and radionuclide activities	Determining and quantifying the sediment sources
Soil and sediment properties raster maps, graphs, tables	Geospatial analysis maps, graphs, and tabular values of the soil and sediment properties	Data and maps will be used for determining spatial correlations between raster maps and feeding erosion process models
FRNs activity concentrations	Data that show the vertical and spatial distribution of the FRNs in the soils	FRN measurement data will be used to estimate soil redistribution rates in the river basins
Soil redistribution rates	The estimation of soil redistribution rates through	Spatial distribution of FRNs inventories in the soil and soil

Dataset name	Description	Data used concerning the research field
(FRN methods)	the application of different conversion models	redistribution rates using geostatistical methods
Spectral indices data	Combination of spectral bands in a multispectral image to show the relative abundance of features of interest	For correlations with other data for detecting and identifying soil erosion
Spatial correlations of raster datasets	Spatial correlations between raster-based models to see the relationship between covariates	For detecting the relationship between soil erosion and soil types, land cover, and terrain properties
Soil erosion vulnerability/risk map (AHP method)	Special-purpose vulnerability maps that show estimated soil erosion based on selected scenarios	For prediction of soil erosion rates and sediment yield at different scenarios for planning appropriate conservation measures
Index of Connectivity map	A map of the sediment delivery model representing sediment movement pathways	The map will help in the determination of the probable sediment sources, movement pathways, and their relocation in the catchment.
WaTEM/SEDEM data	Maps and calculated numeric data on sediment redistribution in the catchment	The data identify erosion and deposition areas within a catchment
Sediment Source Tracing model data (Fingerprinting)	Calculated numeric data and graphs of sediment source contributions modeled with FingerPro	Data will help in detecting the main contributing sources and therefore erosion-endangered areas

Predict-Er project datasets will be available to all project members throughout and after their execution. Upon acceptance of papers by peer-reviewed journals and allocation of DOIs, external researchers will gain complete access to the data for verification and reuse, extending beyond the project's duration. Ensuring transparency and broad utilization, the data will be publicly downloadable from the *Predict-Er* website.

Datasets generated during the project prioritize data security and ethical considerations, abstaining from containing sensitive information. These datasets, formatted in standard types such as DOCX, XLSX, GeoTIFF, JPEG, PDF, etc., comply with interoperability requirements, facilitating seamless integration into diverse platforms.

3.2 Dataset reference and name

To facilitate efficient data discovery and identification, each dataset will include descriptive metadata, encompassing properties such as file name, author, data type, creation date, etc. The overarching goal of the project is to make dataset content easily discernible for all team members and interested researchers. This is achieved through an internal naming convention based on the format: *Predict-Er*-{Work Package}-{File Title}. {File Version}. {File Type Extension} (e.g., *Predict-Er*-WP1-LandCover.01.shp). This standardized approach ensures clarity and accessibility, contributing to the project's overall efficiency and collaborative research efforts.

3.3 Standards and Metadata

i) Microsoft Office 2010 for text-based documents (or any other compatible version) .doc, .docx, .xls, .xlsx, .ppt, .pptx. Also, especially where larger datasets need to be dealt with, .csv and .txt file formats will be used. All finished and approved documents will also be made available as .pdf documents;

ii) Illustrations and graphic design will make use of Excel, PowerPoint, and Photoshop (Format: different types possible, mostly .png), and will be available as .jpg, .psd, .tiff, and .ai files;

iii) PFDs and layouts will preferentially use inkscape.org, an open-source software for vector graphics. (Format: .svg), and will be made available as .png, .jpg, and .pdf files;

iv) MP3 for audio files;

v) Quicktime Movie or Windows Media Video for video files.

We have selected these file formats due to their wide acceptance and common usage. Whenever feasible, files will be transformed into open formats for extended storage purposes.

3.4 Data sharing and access

The mechanisms explained in this section aim at reducing to the maximum the risks related to data storage. Identification of the repository where data will be stored, if already existing and identified, indicating the type of repository. In cases where the dataset cannot be shared, the reasons for this will be stated (e.g. security-related, policies of personal data, ethical, intellectual property, commercial, privacy-related).

Depending on the associated sharing policies, the project's data will undergo diverse curation methods. All non-confidential data, whether open or restricted, will be preserved and made readily accessible to stakeholders throughout and beyond the project's duration. Database compliance measures are in place to enforce security policies, including vulnerability checks and data accuracy verification. The objective is to identify and rectify issues such as excessive user privileges, and weak passwords (or lack thereof), and conduct comprehensive database analyses. To ensure proper data management, the following measures will be implemented: i) Establishing minimum storage requirements to mitigate potential risks, and ii) Implementing access control lists for user and data authentication.

Project data will be centralized and managed securely on a UNFSM Windows server, ensuring protection against data loss and unauthorized access. In addition to being transferred to the server, datasets will be stored on personal data storage devices such as hard drives or DVDs belonging to project members. This serves as an additional backup method for the data. Furthermore, in adherence to the FAIR principles (Findability, Accessibility, Interoperability, and Reusability), the results datasets will be archived in the Zenodo public repository. Each team member will retain data on personal storage devices, with the assurance that the acquired data does not include sensitive or confidential information, mitigating potential security risks associated with losing personal data carriers. *Predict-Er* members will have access to datasets and the authority to modify and supplement data related to their project responsibilities.

All user activities within UNFSM Windows server shared folders, including dataset access, are monitored for the purpose of identifying potentially harmful user behavior. Furthermore, individuals responsible for safeguarding the stored information will be identified.

The research data will be securely preserved and stored for the long term within the *Predict-Er* intranet, housed both on UNSFM web infrastructure and internal hard discs.

3.5 Data archiving and preservation

Data will be organized into files to guarantee accessibility throughout the project and beyond its completion. Each category of data will have its own designated database. All data produced within *Predict-Er* will be consolidated within a digital database dedicated to the *Predict-Er* initiative.

The datasets will be accessible to all project members both during and after the project's execution. Once papers are accepted by peer-reviewed journals and assigned DOIs, other researchers will have complete access to the data for validation and reuse, either during or after the project's execution. The data will be openly downloadable from the *Predict-Er* website. Importantly, data generated during the project will not include sensitive information.

4 Type of data generated in *Predict-Er*

Various types of data will be collected/generated during the Project.

The objective is to create detailed datasets for a typical study area, intended for direct utilization or as simulated data for refining the *Predict-Er* tool. This goal will be realized through comprehensive data collection on the selected catchment, reviewing available data on catchment degradation, evaluating current options with active stakeholder participation, conducting field studies, and performing soil and sediment sampling alongside physicochemical analyses of collected samples.

The aim is to create and implement the groundbreaking *Predict-Er* tool, designed to forecast both the spatial and temporal patterns of soil erosion and sediment movement. This tool has the potential for global applicability once customized with location-specific information. It involves the incorporation of data from various study sites into a unified tool capable of evaluating and quantifying the scope, intensity, and consequences of land degradation, as well as forecasting soil vulnerability under diverse environmental and managerial conditions.

For acquiring high-resolution data for terrain and land cover modeling, the project will employ the following methodologies: 1. Aerial surveying utilizing Light Detection and Ranging (LiDAR) technology deployed via an unmanned aerial vehicle (UAV or drone), providing centimeter-level resolution for the entire catchment area and millimeter-level resolution for investigating the impact of soil microrelief on surface runoff and sediment yield at fallout radionuclides (FRN) transects and gully sampling sites. Additionally, a UAV-based multispectral camera will be utilized; 2. Ground-based surveying employing terrestrial LiDAR and a 360° camera.

If the participant wants to use relevant information from the case study area but is not the creator of the data (e.g. by obtaining relevant data sets or relevant documentation), then it is a participant responsible for determining the source of the data and assessing whether it is the data set contains personal data or data that otherwise compromises privacy. In such case it is

the participant's responsibility to clear personal data from that dataset and prepare it for further dissemination in an appropriate manner allowed form.

5 Making *Predict-Er* data findable

Predict-Er is anticipated to upload most of the generated and collected data to an openly accessible online research data repository. *Predict-Er* will reach out to the ZENODO repository as its preferred data archive, selected based on its compliance with structure, features, and management standards.

Alternative online research data repositories will be evaluated depending on the types and formats of the generated and collected data. Participants involved in the project will be prompted to explore resources like the Registry of Research Data Repositories, and the Directory of Open Access Repositories (OpenDOAR) for comprehensive listings of repositories that could accommodate *Predict-Er* outputs.

The management principles behind storing and making findable data collected through *Predict-Er* that partners should follow are detailed in the:

- a.- Store and make findable any *Predict-Er* data that can be made openly accessible, either in the ZENODO repository or another online data repository suitable for the type and format of data generated or collected.
- b. The organization, data collection, and optimal formatting will fall under the purview of the relevant task leader.
- c. Each task leader will oversee the deposition of pertinent data in ZENODO or another suitable open-access online repository. Data will be available immediately after publication in peer-reviewed scientific articles or similar publications unless there are valid justifications for maintaining data confidentiality.
- d. Each participant is accountable for maintaining records and documentation related to the generated data, ensuring alignment with accepted standards in their respective fields.

To maximize the impact of *Predict-Er* data, the project will facilitate the sharing results and outcomes. Selected data and results will be shared with the scientific community and other interested parties through publications in scientific journals and presentations at conferences and workshops, as well as through the *Predict-Er* tool. Bibliographic metadata must be in a standard format and must contain the name, acronym and grant number, publication date, length of embargo period if applicable, and permanent identifier.

Data resulting from *Predict-Er* activities, which serves as the basis for scientific publications, should be submitted to the relevant *Predict-Er* task and work package leaders. An information template outlining the required descriptive details, as specified by the work package leader, will be distributed to all participants. Upon receiving the data, the work package leader will review each dataset, promptly requesting any necessary additions or modifications to facilitate its upload to open-access repositories. Task leaders must upload the

dataset no later than 60 days after the original publication date. It is the responsibility of project partners to timely prepare and submit the template to expedite this process.

6 Ethical and Legal Considerations in the *Predict-Er* Project: Ensuring Data Integrity and Confidentiality

This section provides a detailed overview of the ethical and legal considerations relevant to the Predict-Er project, which is funded by the Science Fund of the Republic of Serbia under the PRISMA program, #GRANT No. 7047. The discussion covers key aspects such as obtaining consent for data retention and sharing, protecting the identities of individuals and organizations, and ensuring the secure handling of sensitive data during storage and transfer.

The consortium attaches great importance to compliance with the principles of data protection and research ethics. Strict measures are taken to prevent any misuse of sensitive information and personal data is only shared for specific consortium purposes.

Partners involved in the *Predict-Er* project commit to meeting national and EU standards concerning ethics and data management. This entails adherence to ethical principles outlined in the European Code of Conduct for Research Integrity, as well as compliance with relevant international, EU, and national laws, including EU Directive 95/46/EC.

Personal data is handled with utmost confidentiality, stored, analyzed, and utilized anonymously. Individuals are provided comprehensive information regarding the intended use of their data and are required to provide written consent for its collection for scientific purposes.

Anonymization processes are employed to protect the identities of individuals engaged in the project, ensuring that no individual can be identified. Research findings are published in a manner that prevents any possibility of individual identification.

The entire data protection process, including information dissemination to individuals regarding data protection issues, is meticulously designed to comply with national laws, and overseen by legal experts from project partners.

Data collection, storage, protection, retention, and destruction are managed through the project's intranet system, ensuring comprehensive security measures. Participants are thoroughly briefed on data security protocols, anonymity measures, and data usage, and their involvement in project activities is strictly voluntary.