



TECHNISCHE  
UNIVERSITÄT  
WIEN

## Data management plan (DMP)

# Home Advantage in Professional Football – Research Data Management Project

Home Advantage in Football

Version	Effective date	Description of document/changes
1.0	22/11/2025	First version of the DMP – created for the start of the project

Level of  
distribution



This DMP is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0).

This DMP is not published in a public repository and does not have a DOI.

## FWF Data Management Plan (DMP)

I General Information							
I.1 Administrative information	<p><b>PI:</b> Bartosz Ciesielski, bartosz.ciesielski@student.tuwien.ac.at, TU Wien, ROR: ror.org/04d836q62, Project Leader</p> <p><b>FWF project number:</b></p> <p><b>Internal Project ID:</b></p> <p><b>DMP version:</b> 01, 22.11.2025</p> <p><b>Contributors:</b></p> <p>Tomasz Miksa, tomasz.miksa@tuwien.ac.at, ORCID: 0000-0002-4929-7875, TU Wien, ROR: ror.org/04d836q62, Hosting Institution</p>						
I.2 Data management responsibilities and resources	<p><b>Person responsible for data management and DMP:</b> Bartosz Ciesielski, bartosz.ciesielski@student.tuwien.ac.at, TU Wien, ROR: ror.org/04d836q62</p> <p><b>Co-ordination of data management responsibilities across partners:</b> Bartosz Ciesielski, bartosz.ciesielski@student.tuwien.ac.at, TU Wien, ROR: ror.org/04d836q62</p> <p><b>Resources costed in for RDM:</b> There are no costs dedicated to data management and ensuring that data will be FAIR.</p>						
II Data Characteristics							
II.1 Data description and collection or re-use of existing data	<b>Produced datasets:</b>						
	dataset ID	title	type	format	estimated volume	contains sensitive data	description
	-	-	-	-	-	-	-
	No datasets will be produced						
	<b>Reused datasets:</b>						
	dataset ID	title	source		rights (e.g. license)	contains sensitive data	description

	R1	Professional football match results (Ekstraklasa)	<a href="https://www.football-data.co.uk/poland.php">https://www.football-data.co.uk/poland.php</a>		no	...
	R2	Professional football match results (Premier League)	<a href="https://github.com/jalapic/engsoccerdata/tree/master/data">https://github.com/jalapic/engsoccerdata/tree/master/data</a>		no	...
<b>Methods and software used for data generation and reuse</b>  No new data will be generated. Existing football match datasets will be reused and processed using Python (pandas) and Jupyter Notebooks. Raw data will be stored in /data/raw/ and processed outputs in /data/processed/.						
<b>III Documentation and Data Quality</b>						
<b>III.1 Metadata and documentation</b>	<b>Data organisation, metadata, and documentation:</b>  The project follows a clear folder structure separating raw data (/data/raw/), processed data (/data/processed/), notebooks (/notebooks/), reports (/reports/), and source code (/src/). Raw data will remain unchanged, while all transformations and analyses will be stored as new processed files to ensure traceability. Versioning is handled through Git and GitHub. All changes to data processing scripts, notebooks, and documentation are tracked automatically via commits, providing a complete and transparent version history. File naming will follow a consistent structure that reflects content and processing steps. This setup ensures reproducibility, clear documentation of changes, and long-term maintainability of the research data.  The project will provide descriptive metadata to ensure that the datasets can be easily identified, understood, and reused. Metadata will include: dataset title, source URL, data origin, time period covered, variable descriptions (teams, goals, match results, home/away indicators), file formats, and licensing information. As there are no specific domain metadata standards required for this type of sports statistics data, metadata will be documented at the project level through a detailed README file. Each dataset will include clear naming conventions and accompanying descriptions in the /data/processed/ directory. Information about data provenance, reuse conditions, and processing steps will be provided in the documentation so that users can trace how the raw data were transformed. This ensures transparency, discoverability, and reusability of all research outputs. This will help others to identify, discover and reuse our data.  Additionally, we will provide common metadata such as title, description, or keywords when publishing data in open access repositories. In such a case, we will follow the default template provided by the repository, such as Data Cite Metadata or Dublin Core.  As far as possible, we will use controlled vocabularies for our data to allow inter-disciplinary interoperability and machine-actionability.  All analysis steps are documented in the Jupyter notebook (home_advantage_study.ipynb), including data loading, cleaning, transformations, and visualizations. The repository README describes the project structure, datasets, processing workflow, software used, and instructions for reproducing results. Processed data files and visual outputs are stored in /data/processed/ and /reports/figures/ to enable verification and reuse.					

III.2 Data quality control	<b>Data quality control:</b>  The following data quality checks will be done: peer review of data.												
IV Data Storage, Sharing, and Long-Term Preservation													
IV.1 Data storage and backup during the research process	<b>Storage and backup facilities:</b>  No storage options are specified at the moment.												
	<b>Data security and protection of sensitive data:</b>  We pay strict attention to compliance with the relevant institutional and national data protection policies. At this stage, it is not foreseen to process any sensitive data in the project. If this changes, advice will be sought from the data protection specialist at TU Wien, and the DMP will be updated.												
	Access to data during research:												
	<table><tr><td>dataset ID</td><td>selected project members</td><td>all other project members</td><td>the public</td></tr><tr><td>R1</td><td>writing</td><td>reading only</td><td>no access</td></tr><tr><td>R2</td><td>writing</td><td>reading only</td><td>no access</td></tr></table>	dataset ID	selected project members	all other project members	the public	R1	writing	reading only	no access	R2	writing	reading only	no access
	dataset ID	selected project members	all other project members	the public									
R1	writing	reading only	no access										
R2	writing	reading only	no access										
IV.2 Data sharing and long-term preservation	<b>Data publication and access conditions:</b>  As far as possible, obtained datasets will be published in repositories. Details on access conditions, reuse licenses, reasons for restrictions, etc. are collected in the table below.												
	<table><tr><td>dataset ID</td><td>access conditions</td><td>estimated publication date</td><td>location for publication (repository)</td><td>PID</td><td>license</td></tr><tr><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td></tr></table>	dataset ID	access conditions	estimated publication date	location for publication (repository)	PID	license	-	-	-	-	-	-
	dataset ID	access conditions	estimated publication date	location for publication (repository)	PID	license							
	-	-	-	-	-	-							
No new datasets will be published. All reused datasets remain under the licenses of their original providers.													
Methods or software needed to access and use data: The data are provided in standard tabular formats (.xlsx and .csv), which can be opened with common tools such as Excel, Python (pandas), or R. No specialized software is required. The folder structure and documentation in the repository support straightforward reuse.													

	<p><b>Long-term preservation and deletion of data:</b></p> <table border="1" data-bbox="439 248 2092 434"> <thead> <tr> <th data-bbox="439 248 555 368">dataset ID</th><th data-bbox="555 248 848 368">location for long-term storage</th><th data-bbox="848 248 1106 368">minimum retention period (≥ 10 years)</th><th data-bbox="1106 248 2092 368">foreseeable research uses and/or users</th></tr> </thead> <tbody> <tr> <td data-bbox="439 368 555 434">-</td><td data-bbox="555 368 848 434">-</td><td data-bbox="848 368 1106 434">-</td><td data-bbox="1106 368 2092 434">-</td></tr> </tbody> </table> <p>No new datasets will be published. The project uses only existing openly accessible datasets (R1 and R2), which remain available under their original licenses. No additional repository deposit is required.</p>	dataset ID	location for long-term storage	minimum retention period (≥ 10 years)	foreseeable research uses and/or users	-	-	-	-
dataset ID	location for long-term storage	minimum retention period (≥ 10 years)	foreseeable research uses and/or users						
-	-	-	-						
<b>V Legal and Ethical Aspects</b>									
<b>V.1 Legal aspects</b>	<p><b>Personal data:</b></p> <p>At this stage, it is not foreseen to process any personal data in the project. If this changes, advice will be sought from the data protection specialist at TU Wien, and the DMP will be updated.</p> <p><b>Intellectual property rights and rights of use:</b></p> <p>The following individual(s) hold rights and control access to the project data: Access to all datasets will be controlled solely by the project owner (Bartosz Ciesielski). No additional project members have access rights.</p>								
<b>V.2 Ethical aspects</b>	<p><b>Ethical issues:</b></p> <p>No particular ethical issue is foreseen with the data to be used or produced by the project. This section will be updated if issues arise.</p>								