


Data management plan (DMP)

Predicting the Market Value of Football Players Using Various Factors

FAIR-DS-Football-Valuation

| Version | Effective date | Description of document/changes |
|---------|----------------|--|
| 1.0 | 25/05/2026 | Final version of the DMP - FAIR Data Science course submission, TU Wien 2026 |
| | | |
| | | |
| | | |

| | | |
|-----------------------|---|---|
| Level of distribution |  | This DMP is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). It is publicly available under [DMP DOI to be added after T4.4 deposit is published on TUWRD]. |
|-----------------------|---|---|



TECHNISCHE
UNIVERSITÄT
WIEN

| | | |
|--|--|--|
| | | [If you do not publish this DMP, remove the CC BY icon and change the text above accordingly.] |
|--|--|--|

FWF Data Management Plan (DMP)

This DMP follows the official FWF template. When revising this DMP, please check the guidance from FWF for more details: [FWF DMP - Guidance and Template](#)

| I General Information | | | | | | | | | | | | | | | | | | | | |
|--|---|-----------------|--------|------------------|-------------------------|--|------------|-------|------|--------|------------------|-------------------------|-------------|----|--------------------------------------|-----------------|-----|---------|----|--|
| I.1 Administrative information | <p>FWF project number: N/A – TU Wien FAIR Data Science course exercise (2026)</p> <p>Internal Project ID: FAIR-DS-Football-Valuation</p> <p>DMP version: 1.0, 25.05.2026</p> <p>Contributors:</p> <p>Muhammad Bilal Hussain, bilal.hussain3223@gmail.com, ORCID: 0009-0000-2512-9167, Data Manager</p> <p>Edeh Ekene, edehjamesraphael@gmail.com, ORCID: 0009-0007-2481-389X, Data Manager</p> <p>Muhammad Athar Riaz, bilal.hussian2332@gmail.com, ORCID: 0009-0004-3337-8672, Data Manager</p> | | | | | | | | | | | | | | | | | | | |
| I.2 Data management responsibilities and resources | <p>Person responsible for data management and DMP: Edeh Ekene, edehjamesraphael@gmail.com, ORCID: 0009-0007-2481-389X</p> <p>Co-ordination of data management responsibilities across partners: Shared across all four group members. Each member is responsible for their assigned work packages as documented in the repository contribution table.</p> <p>Resources costed in for RDM: 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 | <p>Produced datasets:</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 10%;">dataset ID</th> <th style="width: 20%;">title</th> <th style="width: 10%;">type</th> <th style="width: 10%;">format</th> <th style="width: 10%;">estimated volume</th> <th style="width: 10%;">contains sensitive data</th> <th style="width: 30%;">description</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>Football Player Market Value Dataset</td> <td>Structured text</td> <td>csv</td> <td>< 50 MB</td> <td>no</td> <td>Player valuation predictions (CSV), trained XGBoost regression</td> </tr> </tbody> </table> | | | | | | dataset ID | title | type | format | estimated volume | contains sensitive data | description | P1 | Football Player Market Value Dataset | Structured text | csv | < 50 MB | no | Player valuation predictions (CSV), trained XGBoost regression |
| dataset ID | title | type | format | estimated volume | contains sensitive data | description | | | | | | | | | | | | | | |
| P1 | Football Player Market Value Dataset | Structured text | csv | < 50 MB | no | Player valuation predictions (CSV), trained XGBoost regression | | | | | | | | | | | | | | |

| IV Data Storage, Sharing, and Long-Term Preservation | | | | | | | | | | | | | |
|--|---|----------------------------|---------------------------------------|--|--|-----|---------|--------------|-----------|---------|-----------------------|--|--|
| <p>IV.1 Data storage and backup during the research process</p> | <p>Storage and backup facilities:</p> <p>For the duration of the project, storage and backup of data will be ensured by Edeh Ekene (acting as the person responsible for data management and DMP) in cooperation with the system operator. Input data is stored on TU Wien DBRepo (test instance) as a structured relational database, accessible at https://test.dbrepo.tuwien.ac.at/database/598ce585-d8b5-4a97-8f19-cb085d4a5b1e. Source code, metadata, documentation, and version control are managed via a public GitHub repository. Generated outputs and trained models are published on TU Wien Research Data Repository (TUWRD) and Zenodo. P1 (Football Player Market Value Dataset) will be stored on GitHub Repository. External storage will be used because public github repository used for source code, metadata, documentation, and version control.</p> <p>Data security and protection of sensitive data:</p> <p>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.</p> <p>Access to data during research:</p> <table border="1" data-bbox="439 727 1644 884"> <thead> <tr> <th>dataset ID</th> <th>selected project members</th> <th>all other project members</th> <th>the public</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>writing</td> <td>reading only</td> <td>no access</td> </tr> </tbody> </table> <p>All incidents will be handled individually by an incident response team that is maintaining the affected service.</p> | dataset ID | selected project members | all other project members | the public | P1 | writing | reading only | no access | | | | |
| dataset ID | selected project members | all other project members | the public | | | | | | | | | | |
| P1 | writing | reading only | no access | | | | | | | | | | |
| <p>IV.2 Data sharing and long-term preservation</p> | <p>Data publication and access conditions:</p> <p>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.</p> <table border="1" data-bbox="439 1098 2092 1374"> <thead> <tr> <th>dataset ID</th> <th>access conditions</th> <th>estimated publication date</th> <th>location for publication (repository)</th> <th>PID</th> <th>license</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>Open</td> <td>2026-05</td> <td>TU Wien Research Data</td> <td>https://doi.org/10.70124/g7pw4-cd077 (model); https://doi.org/10.5281/zenodo.20377147 (generated data); https://doi.org/10.5281/zenodo.20377147</td> <td>CC-BY-4.0</td> </tr> </tbody> </table> | dataset ID | access conditions | estimated publication date | location for publication (repository) | PID | license | P1 | Open | 2026-05 | TU Wien Research Data | https://doi.org/10.70124/g7pw4-cd077 (model); https://doi.org/10.5281/zenodo.20377147 (generated data); https://doi.org/10.5281/zenodo.20377147 | CC-BY-4.0 |
| dataset ID | access conditions | estimated publication date | location for publication (repository) | PID | license | | | | | | | | |
| P1 | Open | 2026-05 | TU Wien Research Data | https://doi.org/10.70124/g7pw4-cd077 (model); https://doi.org/10.5281/zenodo.20377147 (generated data); https://doi.org/10.5281/zenodo.20377147 | CC-BY-4.0 | | | | | | | | |

| | | | | | |
|--|--|--|--|----------------------|--|
| | | | | nodo.20357906 (code) | |
|--|--|--|--|----------------------|--|

Repository description:

TU Wien Research Data is an institutional repository of TU Wien to enable storing, sharing and publishing of digital objects, in particular research data. It facilitates the funders' requirements for open access to research data and the FAIR principles by making research output findable, accessible, interoperable, and reusable. A DOI is assigned to each dataset published in TU Wien Research Data. This service is developed by the TU Wien Center for Research Data Management and hosted by TU.it. <https://researchdata.tuwien.ac.at/>

Methods or software needed to access and use data: Potential users may require common data science and machine learning tools such as Python, Jupyter Notebook, pandas, scikit-learn, XGBoost, and spreadsheet software capable of opening CSV or XLSX files. The repository is hosted on GitHub and can be accessed using standard web browsers and Git-based version control tools.

Long-term preservation and deletion of data:

| dataset ID | location for long-term storage | minimum retention period (≥ 10 years) | foreseeable research uses and/or users |
|------------|--------------------------------|---------------------------------------|--|
| P1 | TU Wien Research Data | 10 years | <p>The target audience includes researchers, data scientists, sports analytics researchers, football analysts, and machine learning practitioners interested in football player valuation and sports data analysis.</p> <p>The datasets and generated outputs may also be useful for educational purposes, reproducible machine learning research, FAIR data management studies, and analytical experimentation involving regression models and player performance evaluation.</p> |

V Legal and Ethical Aspects

V.1 Legal aspects

Personal data:

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.

Intellectual property rights and rights of use:

| | |
|-----------------------------------|--|
| | <p>The following individual(s) hold rights and control access to the project data: All project contributors will have controlled access to the datasets and repository resources through GitHub and project management platforms. Edeh Ekene (D) is responsible for repository releases, DOI integration, DBRepo infrastructure, API reimplementation, and data management plan finalisation. Muhammad Bilal Hussain (C) is responsible for repository setup, README documentation, data normalisation, DBRepo data loading, and FAIR4ML metadata. Muhammad Athar Riaz (B) is responsible for semantic mappings, CodeMeta integration, maDMP finalisation, and generated-data deposits. Edeh Ekene (D) is responsible for repository releases, DOI integration, and preservation-related tasks. Public access will be provided only to the final FAIR-compliant repository, metadata artefacts, documentation, and permitted generated outputs after project completion.</p> |
| <p>V.2 Ethical aspects</p> | <p>Ethical issues: 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> |