TECHNISCHE UNIVERSITÄT WIEN

Data management plan (DMP)

Impact of highly rated game releases on Activision Stock Price

EOSCSecretariat.eu

Version	Effective date	Description of document/changes
1.0	10/05/2023	First version of DMP – created for start of project (deliverable 13.05.2023)
2.0	11/05/2023	Second version of DMP – prepared for midterm review (deliverable 13.05.2023)

Level of distribution		This DMP is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).
		DOI: https://doi.org/10.5281/zenodo.7930435

Project details

Project Coordinator Principal Investigator	Dominik Granig, e11708471@student.tuwien.at.at, ORCID iD: 0009-0009-8898-5361
Contact person (responsible for data management and DMP)	Dominik Granig, e11708471@student.tuwien.at.at, ORCID iD: 0009-0009-8898-5361
Contributors	Tuvshin Selenge, ORCID iD: 0009-0001-3293-2959, TU Wien, Data Manager Tuvshin Selenge, ORCID iD: 0009-0001-3293-2959, TU Wien, Data Collector
Start date	2023-04-30
End date	2023-05-14
Funder, funding programme, grant number	
Internal project number TU Wien	

List of acronyms

DMP	data management plan
RDM	research data management
…	…
…	…
…	…
…	…
…	…
…	…

Content

1. Introduction

Science Europe practical guide, FAIR data

A DMP is a structured document that keeps record of what research data is created and what happens to that data during and after a project. It helps with planning the research process and defining responsibilities in a research project involving several researchers or institutions.

For writing this DMP, we followed the recommendations of Science Europe as they reflect the guidelines agreed upon by the major funders in Europe.

To make our data FAIR, they generally will be treated according to the following criteria:

- We will make our data findable, by uploading it to a data repository that provides a persistent identifier, and adding relevant metadata.
- We will make our data accessible by providing open access to data, wherever possible. In cases, where open access is not possible, we will provide meaningful metadata plus contact information for access requests.
- We will make our data interoperable by providing and describing data in a way that is common within our domain by using the same file formats, schemas and vocabularies. We will provide good documentation for all our datasets.
- We will make our data reusable by adding metadata and comprehensive Readme files to all published datasets. The descriptions include details on the methodology used, analytical and procedural information. In case of publication, licenses for code and data will always be assigned and clearly marked.

Relevant Policies and Guidelines

- TU Wien Policy for Research Data Management: https://www.tuwien.at/index.php?eID=dms&s=4&path=Directives%20and%20Regulations%20of%20the%20Rectorate/Policy%20for%20Research%20Data%20Management.pdf
- TU Wien Code of Conduct – Rules to Ensure Good Scientific Practice: https://www.tuwien.at/index.php?eID=dms&s=4&path=Directives%20and%20Regulations%20of%20the%20Rectorate/Code%20of%20Conduct%20%E2%80%93%20Rules%20to%20Ensure%20Good%20Scientific%20Practice.pd
- Directives and Regulations of the TU Wien Rectorate: https://www.tuwien.at/en/tu-wien/organisation/central-divisions/data-protection-and-document-management/directives-regulations/
- TU Wien Data Protection: https://www.tuwien.at/en/tu-wien/organisation/central-divisions/data-protection-and-document-management/data-protection-at-tu-wien
- European Commission's document on Ethics and Data Protection: https://ec.europa.eu/info/sites/info/files/5._h2020_ethics_and_data_protection.pdf
- Other (e.g. from a project partner)

1. Data description

1a Lists of datasets that will be reused or produced

Produced datasets

dataset ID	title	type	format	estimated volume	contains sensitive data
P1	yFinance Historical Stock Price	Structured text		100 MB	no

Reused datasets

dataset ID	title	PID (e.g. DOI) or source	rights (e.g. license)	contains sensitive data
R1	Steam_games	10.5281/zenodo.7904761	CC0 1.0 Universal	no
R2	Activision stock price	10.5281/zenodo.7904772	CC0 1.0 Universal	no

1b Data generation and reuse

Methods and software used for data generation and reuse

The project involved gathering data from the Steam Store and SteamSpy APIs to create a comprehensive dataset on games available on the Steam platform. This dataset provides valuable information about various aspects of games, such as their genres and the estimated number of owners. The data was collected around May 2019, including most games released prior to that date. Non-games and unreleased titles were filtered out, resulting in a dataset suitable for analysis. To complement the game data, historical price data for Activision was obtained.

The dataset included price data from 2010 onwards, collected by a Kaggle user using an open-source API. The licensing terms specified the usage of the Reddit API. In order to calculate ratings for the Activision games, a rating formula provided by SteamDB, a trusted third-party provider for Steam data, was implemented. This formula was utilized to assess the quality and reception of games. The formula was encapsulated in a Python method named GetRating, which takes the number of positive votes and negative votes as inputs. By applying the formula, the GetRating method calculates a rating score for each game. The GetRating method calculates the rating score based on the provided number of positive votes and negative votes for a game. The formula considers the average of the votes and the total number of votes, adjusting the score based on the logarithm of the total votes. The final rating score is obtained by multiplying the calculated score by 100.

By utilizing the GetRating method, the analysis was able to assign ratings to the Activision games based on the number of positive and negative votes they received. This allowed for a quantitative assessment of the game's reception among users and provided insights into the relationship between game ratings and the stock price performance of Activision.

Throughout the project, the analysis and outputs were coded in Python, leveraging various packages such as matplotlib, pandas, datetime, and numpy. These packages facilitated data manipulation, visualization, and time series analysis to gain insights into the impact of highly-rated game releases on the Activision stock market.

2. Documentation and data quality

2a Data organisation, metadata and documentation

The dataset is structured in CSV format, and its creation timestamp is included as part of the project documentation, which is accessible on GitHub. Version control is enabled because the project follows a structured approach where each dataset is assigned a Digital Object Identifier (DOI), which Zenodo utilizes to provide access to all versions of the dataset.

As there are no domain-specific metadata standards applicable, we will provide a README file with an explanation of all values and terms used at dataset level. 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.

A far as possible, we will use controlled vocabularies for our data to allow inter-disciplinary interoperability and machine-actionability.

To ensure the accuracy of the price data, we conducted validation by comparing it with other reliable financial sources. Similarly, the data on video games was validated by leveraging the Steam API, which is a reputable platform widely used for game distribution. In order to promote data reuse and accessibility, we will make all the data available in our GitHub repository. This will include the dataset files as well as a detailed README file that serves as comprehensive documentation, providing information about the dataset structure, variables, and any necessary instructions or explanations.

2b Data quality control

The following data quality checks will be done: calibration, and peer review of data.

3. Storage and backup during research process

3a Storage and backup facilities

For the duration of the project, storage and backup of data will be ensured by the project manager in cooperation with the responsible representative of TUgitLab. The infrastructure of TU Wien will be used for this purpose.

R1 (Steam_games), R2 (Activision stock price) will be stored in TUgitLab: TUgitLab is a central and readily available network drive with daily backups and regular snapshots provided by TU.it.

It is suitable for storing data with moderate access requirements, but high availability demands that allows full control of allocating authorisations. Only authorized staff members will have access.

3b Data security and protection of sensitive data

We pay strict attention to compliance with the relevant institutional and national data protection policies listed in the introduction of this document. 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 (Verena Dolovai), and the DMP will be updated.

Access to data during research:

dataset ID	selected project members	all other project members	the public
P1	writing	writing	writing
R1	writing	writing	writing
R2	writing	writing	writing

All incidents will be handled individually by an incident response team that is maintaining the affected service.

4. Legal and ethical requirements

4a 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 (Verena Dolovai), and the DMP will be updated.

4b Intellectual property rights and ownership

There are no legal restrictions on the processing and disclosure of our data.

4c Ethical issues

No ethical issue is foreseen with the data to be used or produced by the project. This section will be updated if issues arise.

5. Data sharing and long-term preservation

5a Data publication and access conditions

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.

dataset ID	access conditions	restrictions / embargo reasons	estimated publication date	location for publication (repository)	PID	license
P1	Open		2023-03-14	TU Wien Research Data		CC BY 4.0

The repositories used in this project are described in the following paragraph:

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 re-usable. This service is developed by the TU Wien Center for Research Data Management and hosted by TU.it. https://researchdata.tuwien.at/

Methods or software needed to access and use data: In order to replicate the analysis and utilize the data, potential users will need to have Python installed as their programming language. Python is required to run the analysis code and perform any necessary modifications or extensions. The data can be accessed through my GitHub repository, where users can find all the necessary files, including the datasets and README documentation. Alternatively, users can also retrieve the data using the Digital Object Identifier (DOI) associated with the project, which provides a direct link to the data repository.

5b 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	The analysis and outcomes of this project are made accessible to students and the general public. By providing the necessary documentation, datasets, and methodologies, individuals can examine the analysis process and draw their own conclusions regarding the impact of game releases on the stock prices of publishers or developers. This empowers students and the general public to form their own opinions on the topic and encourages critical thinking and further exploration of the relationship between game releases and stock market performance

6. RDM responsibilities and resources

6a RDM-roles and responsibilities

The PI/data officer will direct the data management process overall, with the research assistant responsible for ensuring metadata production, day-to-day cross-checks, back-up and other quality control activities are maintained. The lead country researchers will be responsible for routine supervision of the dataset development.

6b Resources

The costs dedicated to data management and ensuring that data will be FAIR:

Cost name	Cost type	Description	Unit	Value
Data Manager	Personnel	Hiring dedicated data managers.	1	Data Manger 50.000 € per Year
Personnel costs	Personnel	Hiring dedicated research assistant.	1	Research Assistant 30.000 € per Year
Estimated total costs				**100.000,00 EUR**