1. It is often possible to publish metadata about research data even if the actual data cannot be opened.

- What types of information and documentation would be useful for you to include with your data or software code to ensure its findability and interpretability?

- My work will consist of data from games, alongside modeling scripts for predictive analytics. The basic metadata I can publish will be title and description of the dataset without sharing the data, keywords or hashtags, information about me or ORCID ID, and time period and location of collected data. For a more detailed one, I could include variables, definitions and units of the dataset, the structure of the folders, methodology for data collection and processing, and licensing. Including DOI would also be important. For my software, it will be similar but slightly different. First of all, a README file would be essential to tell what the code is about and how it is, which will contain anything from description, aim, research, to the explanation of files and libraries. It would also be good to include example data structure, but also on limitations of the code. In my case, besides the basic things, it would also be good idea to put modeling notes on predictive AI, versions and cleaning scripts.

- What benefits do you think would come from publishing this information?

- Sharing the metadata will make the research more transparent and easier to understand. It will also keep the data more organised and easier to categorise and search for other people who are interested. FAIR principles emphasises that research should be discoverable through identifiers, keywords, and descriptions that are clear. I agree to this as well. I would rather use and cite research that others can find through online or library, not one that is hard to find and only I know about. Another benefit that I will have is organisability for myself, since it might be hard for me to track what I did without these informations published, like versions of code for example.

2. Find a data repository in your field using the website https://www.re3data.org/. Evaluate the repository's reliability and consider whether your research data would be suitable for that archive. Justify your response.

- At first, I started with searching data analytics in the field, with the filter being in the EU. There was only one and it was called EUDAT. I haven't heard about it, but upon further research, I found that it's a collection of services that support data storage, sharing and long term preservation, similar to CSC (Or IDA). The difference seems like that EUDAT is more focused on transnational projects, whereas CSC is more integrated into Finland's ecosystem. Naturally as a Finnish university student, it would be best if I used CSC (IDA), but it's not bad to consider EUDAT or other repositories like Zenodo for international research access possibilities. I mentioned Zenodo because it's one of the major candidate in EU, known for CERN to fund it. It supports DOI, backup and versioning as well. Less common but perhaps worth mentioning is Github for code

development. It’s one of the major free and accessible platform for coders to upload their code or works and work together. Although it’s not a place for long term preservation data repository, it can be connected to Zenodo so Zenodo archives releases from github, and it assigns DOI to specific releases.

3. What aspects related to data security, privacy, or research ethics should be taken into account with your project's research data?

- Since my research touches on sensitive behavioural data, the primary and main concern is privacy and confidentiality. As we learned in the lecture, personal data must be protected at all stages, and confidentiality is needed when a dataset contains personally identifiable information. So anonymisation is for sure needed. Things such as player ID, demographic info, or any other information should be blurred or protected through rigorous encryption, or maybe not included even if it compromises reliability of study. If personal data is collected or used, I should use secure environments like SD desktop through CSC or university, and make sure to get consent forms. Also, the personal data storage time is of the concern, and it should be deleted within agreed time in the consent form or 2 years, if I’m not mistaken, following the data deletion policy.