

7 DATA MANAGEMENT PLAN

Introduction

This Data Management Plan (DMP) covers the data which will be collected under the proposed project entitled “Notoriously Toxic” and is designed to be consistent with the NSF Policy on Dissemination and Sharing of Research Results. In accordance with this policy, this plan does not include preliminary analyses (including raw data), drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. Furthermore, data may be temporarily withheld from distribution and other proposed data management to protect intellectual property or abide by IRB or NDA guidelines. This plan will make certain that the data produced during the period of this project is appropriately managed to ensure its usability, access, preservation, and confidentiality.

Description of Data

The types of data to be produced in the course of the project include, (1) chat logs from multiplayer online games, and (2) the code used to collect, clean, and analyze those logs. Non-redacted data may include usernames, timestamps, and free text fields. Redacted data may include timestamps, key anonymized usernames, and scrubbed text fields. The project will not explicitly collect chat log data, but will encourage participants to share data pursuant to their IRB and NDA guidelines.

Standards for Data and Metadata Format and Content

For preservation and long-term access, data collection will be accompanied with proper documentation and associated metadata. Files will include the data itself, documentation files which a description of how the data was collected, and metadata in the form of processed logs.

Data Access, Sharing and Archiving

The researchers associated with this study are not aware of any reasons that might prohibit the sharing of the data to be generated under this project for public use and potential secondary uses, assuming data is handled consistent with IRB and NDA guidelines. The principal investigators retain the right for first use of the data. The principal investigators will house data generated from this project on servers at their institutions. Anonymized data will be available for access and sharing as soon as is reasonably possible, normally not longer than one year after publication of the data. The raw data from the proposed research will be archived locally by the researchers on regularly backed-up computers in encrypted formats, and will be preserved for at least three years beyond the award period. Prior to the workshop, participants will be encouraged to share either complete or subsets of their research data amongst the group under proper non-disclosure agreements.

Data Confidentiality

Research records will be kept confidential, and access will be limited to the PI, primary research team members, and project participants. Data will be housed on a local server controlled by the PI, and will be accessible via SSH and VPN. Data containing identifiable information, or information covered by an NDA, will be held in an encrypted format.

Intellectual Property and Sharing of Research Resources

Intellectual property and data generated under this project will be administered in accordance with University and NSF policies. Materials generated under the project will be disseminated in accordance with University/Participating institutional and NSF policies.

Access to databases and associated software tools generated under the project will be available for educational, research and non-profit purposes. Such access will be provided using web-based applications, as appropriate.

Publication of data shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Published research data will be redacted to prevent the disclosure of personal identifiers.