

0. Proposal name
<i>Explaining Time Series Downsampling</i>
1. Description of the data
<p>1.1 Type of study</p> <p><i>Improving how data practitioners better understand and explain the impact of downsampling time series data, this study includes time series data sets, exploratory analysis, of these datasets in R, user research and comparative surveys of compression algorithms, common evaluation metrics, and time series visualisations.</i></p> <p>1.2 Types of data</p> <p><i>Both quantitative data from open-source data sets and qualitative data from user research will be used in this study.</i></p> <p><i>The quantitative data sets are being sourced from the Alan Turing Institute 'AnnotateChange' (https://github.com/alan-turing-institute/AnnotateChange) and 'Turing Change Point Dataset' (https://github.com/alan-turing-institute/TCPD/tree/master). The 'AnnotateChange' repository was created to collect annotations of time series data to construct the 'Turing Change Point Dataset' repository by Van den Burgh and Williams (2020). The publicly available data in these datasets will be used within the stipulated licensing agreement(s) stipulated by the data owner(s).</i></p> <p><i>The qualitative data will be collected from interviews with decision-makers and practitioners..</i></p> <p>1.3 Format and scale of the data</p> <p><i>The qualitative data will be collected via recorded video calls with participants and notes taken in csv files.</i></p> <p><i>The quantitative demo data from 'AnnotateChange' will be used initially as JSON scripts are provided. A subset of data is likely to be selected from the 'Turing Change Point Dataset', which includes 37 datasets of time series data from a variety of contexts as well as 5 quality control datasets. The data across these repositories are provided in different formats pending where the original data is hosted. The FAIR principles (Findability, Accessibility, Interoperability and Reusability) are satisfied by the Alan Turin Institute's use of these datasets as benchmark suites. The project will also utilise R, RStudio and a variety of R packages, like RMarkdown to ensure reproducibility.</i></p>
2. Data collection / generation
<p>2.1 Methodologies for data collection / generation</p> <p><i>No new time series data will be collected for this study. The results created during this research will be clearly documented in the report, tables and reproducible code housed in GitHub (https://github.com/MoFrod/downsampling_timeseries/tree/main).</i></p> <p><i>New data will be collected from interviews with data practitioners and decision-makers considering time series data. There is no data source that currently sets out the perspectives of these users, so new data is needed. This new qualitative data will be collected from a standardised list of interview questions that are asked to volunteer participants.</i></p> <p>2.2 Data quality and standards</p> <p><i>The quantitative data quality of the datasets is acceptable in line with the FAIR principles, and data cleaning will be conducted as required. The qualitative data quality cannot be determined until it is collected.</i></p>
3. Data management, documentation and curation
<p>3.1 Managing, storing and curating data.</p> <p><i>The qualitative data generated by user research interviews will be stored in password protected</i></p>

folders as video recordings of the interviews. Notes taken from these interviews will be stored as password protected csv files.

The quantitative data will be stored locally with instructions of how to access the data specified in the project GitHub repository included in 2.1.

3.2 Metadata standards and data documentation

Descriptions of the data used throughout this study will be provided as part of the final report and any details needed for reproducibility will be documented in README files in the project GitHub repository shared in 2.1.

The qualitative data will not be shared beyond the researcher and supervisor in line with the participants' wishes outlined by completed consent forms.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

Not applicable.

4.2 Main risks to data security

The data security risks for this study are in-relation to personal or sensitive data that may be collected as part of the user research. To mitigate these risks, participants will be required to complete a consent form before participating, all information will be anonymized in a password protected file saved in a separate location to the interviews, and participants will be reminded that the session is being recorded before recording starts. Recorded interviews will be stored in password protected folders as video recordings of the interviews. Notes taken from these interviews will be stored as password protected csv files.

5. Data sharing and access

5.1 Suitability for sharing

The quantitative data used or generated in this project is suitable for sharing as the original data is already publicly available and cited in a published paper. Thew qualitative data will not be suitable for sharing outside anonymized quotes used for the final report.

5.2 Discovery by potential users of the research data

The quantitative data being used for this project is already publicly available. The code that generates results from this data will be pushed in the GitHub repository shared in section 2.1. If the project generates useful insights, the researcher will consider publication to help achieve the research aim.

5.3 Governance of access

The qualitative data generated in this study will not be shared beyond the researcher and supervisor in line with the consent form.

5.4 The study team's exclusive use of the data

Not applicable.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

The raw qualitative data from recorded interviews cannot be shared. However, two options have been included in the consent form to help facilitate sharing of the anonymised data for publication and further research. Sharing of anonmyised and aggregated data may be permitted if the participants consent.

5.6 Regulation of responsibilities of users

If anonymised and aggregated data can be shared, it may only be shared for further research on the topic outlined by the consent form. This will be specified clearly as appropriate.

6. Responsibilities

The Dissertation Supervisor, Matthew Forshaw, will be supporting this project . He has the

<i>same responsibilities as the research and also has the responsibility of oversight.</i>	
7. Relevant institutional, departmental or study policies on data sharing and data security	
Policy	URL or Reference
Data Management Policy & Procedures	https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf
Data Security Policy	https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2_1%20SJ%20v0.1%20amended%202022-08-05.pdf
Data Sharing Policy	https://www.ncl.ac.uk/media/wwwnclacuk/freedomofinformation/dataprotection/Data%20Protection%20Policy%20June2020.pdf
8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details	
Morgan Frodsham	