Plan Overview

A Data Management Plan created using DMPonline

Title: Practical and reliable Federated Learning in the wild

Creator: Bart Cox

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2021)

ID: 114051

Start date: 01-12-2021

End date: 01-12-2025

Last modified: 13-01-2023

Copyright information:
The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Practical and reliable Federated Learning in the wild

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

None, I've followed the RDM course.

2. Date of consultation with support staff.

2023-01-13

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Public datasets (Vision, Language datasets). Examples are MNIST, Cifar-10, CelebA	Binary	Available online in the public domain or free of use with citation.	Input for experiments	Github	Public open data
System performance traces. The machines used run in a cluster.	csv	Generated from experiments	To evaluate system performance.	Local machines and Github	PhD student and supervisor
Logs	csv	Generated by experiments	To evaluate system performance and Machine Learning performance.	Local machines and Github	PhD student and supervisor
Experiment configuration	yaml	Created by PhD student	Input for experiment setup	Local machines and Github	PhD student and supervisor

4. How much data storage will you require during the project lifetime?

- 250 GB - 5 TB

II. Documentation and data quality

5. What documentation will accompany data?

- README file or other documentation explaining how data is organised

III. Storage and backup during research process

6. Where will the data (and code, if applicable) be stored and backed-up during the project lifetime?

- Another storage system - please explain below, including provided security measures

Projects are stored on Github.
As much as possible is stored as open-source in public repositories.

IV. Legal and ethical requirements, codes of conduct

7. Does your research involve human subjects or 3rd party datasets collected from human participants?

- No

8A. Will you work with personal data? (information about an identified or identifiable natural person)

If you are not sure which option to select, ask your [Faculty Data Steward](#) for advice. You can also check with the [privacy website](#) or contact the privacy team: privacy-tud@tudelft.nl

- No

The data sets used for this research all live in the public domain and/or are commonly used in Machine Learning papers/research.

8B. Will you work with any other types of confidential or classified data or code as listed below? (tick all that apply)

If you are not sure which option to select, ask your [Faculty Data Steward](#) for advice.

- No, I will not work with any confidential or classified data/code

The data sets used for this research all live in the public domain and/or are commonly used in Machine Learning papers/research.

9. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your [Faculty Contract Manager](#) when answering this question. If this is not the case, you can use the example below.

The data sets used for this research all live in the public domain and/or are commonly used in Machine Learning papers/research.
The ownership of the data sets are managed by the authors that originally published the dataset.
In this research will we refer to the original publication.

V. Data sharing and long-term preservation

26. What data will be publicly shared?

- All data (and code) produced in the project
- All data (and code) underlying published articles / reports / theses
- All validated non-positive results

The data and code is publicly shared as soon as possible (most likely after an accepted publication).

28. How will you share your research data (and code)?

- I will upload the data to another data repository (please provide details below)

Data and code are (eventually) publicly available on Github.
The repository used for this research can be found at https://github.com/bacox/fltk.

30. How much of your data will be shared in a research data repository?

- < 100 GB

31. When will the data (or code) be shared?

- As soon as corresponding results (papers, theses, reports) are published

32. Under what licence will be the data/code released?

- BSD

Choice of license is influence by the use of third-party code included in the project.

VI. Data management responsibilities and resources

33. Is TU Delft the lead institution for this project?

- Yes, the only institution involved

34. If you leave TU Delft (or are unavailable), who is going to be responsible for the data resulting from this project?

Professor Jérémie Decouchant J.Decouchant@tudelft.nl

35. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Open-source projects are free of costs when hosted on Github.
A copy of the code and data can be published on the 4TU.ResearchData.
4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. We do not expect to exceed this and therefore there are no additional costs of long term preservation.