Horizon 2020

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

What types and formats of data will the project generate/collect?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): There are three .csv files containing training data, test data and solutions for the test data.

Dataset Source Code: The source code is split between .py text files which are standard Python script files and .ipynb Jupyter Notebook files (which are in principle special JSON files).

Dataset Result files: There are several .csv files containing row IDs and classification results,

Will you re-use any existing data and how?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Re-used

Dataset Source Code: Created

Dataset Result files: Created

What is the origin of the data?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Mentioned at the provided URLs

Dataset Source Code:

Dataset Result files:

What is the expected size of the data?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

To whom might it be useful ('data utility')?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Since the data produced in this experiment is for the sake of learning to write a DMP only, none if it has any long-term value except maybe the DMP itself, which could be useful for future reference for the author himself and maybe other students writing their first DMP.

Dataset Source Code:

Dataset Result files:

2. FAIR data

2.1 Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Yes Handle / DOI

Dataset Source Code: Yes Handle / DOI

Dataset Result files: Yes Handle / DOI

What naming conventions do you follow?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Not yet

Dataset Source Code:

Dataset Result files:

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): The dataset is static.

Dataset Source Code: The source code is under git version control and uploaded to a central GitHub repository.

Dataset Result files: The source code is under git version control and uploaded to a central GitHub repository.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

- automatic:
- semi-automatic:
- manual:

Dataset Source Code:

- automatic:
- semi-automatic:
- manual:

Dataset Result files:

- automatic:
- semi-automatic:
- manual:

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain data sets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Yes, externally for everyone

Dataset Source Code: Yes, externally for everyone

Dataset Result files: Yes, externally for everyone

How will the data be made accessible (e.g. by deposition in a repository)?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Generic data center: Zenodo, GitHub

Dataset Source Code: Generic data center: Zenodo, GitHub

Dataset Result files: Generic data center: Zenodo, GitHub

What methods or software tools are needed to access the data?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

Is documentation about the software needed to access the data included?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

Is it possible to include the relevant software (e.g. in open source code)?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Yes, externally for everyone

Dataset Source Code: Yes, externally for everyone

Dataset Result files: Yes, externally for everyone

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Generic data center: Zenodo, GitHub (not certified)

Dataset Source Code: Generic data center: Zenodo, GitHub (not certified)

Dataset Result files: Generic data center: Zenodo, GitHub (not certified)

Have you explored appropriate arrangements with the identified repository?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine-readable license)?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

How will the identity of the person accessing the data be ascertained?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): The local machine is secured by a password and the harddrive is locked inside an apartment, but the GitHub repository will be publically available for viewing and forking. Pushing changes can only be done by collaborators, whose accounts are secured by passwords as well.

Dataset Source Code:

Dataset Result files:

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

- No fixed system for the description is used

Dataset Source Code:

Dataset Result files:

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Other: MIT

Dataset Source Code: Other: MIT

Dataset Result files: Other: MIT

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Dataset Pre-existing Data (Breast cancer & Arrhythmia): April 22, 2019

Dataset Source Code: April 22, 2019

Dataset Result files: April 22, 2019

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Since the data produced in this experiment is for the sake of learning to write a DMP only, none if it has any long-term value except maybe the DMP itself, which could be useful for future reference for the author himself and maybe other students writing their first DMP.

Dataset Source Code:

Dataset Result files:

How long is it intended that the data remains re-usable?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): For a minimum of one year

Dataset Source Code: For a minimum of one year

Dataset Result files: For a minimum of one year

3. Allocation of resources

What are the costs for making data FAIR in your project?

Personnel:

Other:

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

-

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

Preservation decisions

Costs

Personnel: 0 PM

Other:

4. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

Dataset Pre-existing Data (Breast cancer & Arrhythmia): Generic data center: Zenodo, GitHub (not certified)

Dataset Source Code: Generic data center: Zenodo, GitHub (not certified)

Dataset Result files: Generic data center: Zenodo, GitHub (not certified)

5. Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

Dataset Source Code:

Dataset Result files:

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

Dataset Pre-existing Data (Breast cancer & Arrhythmia):

- Personal data: No
- Informed consent: The "informed consent" is not obtained

Dataset Source Code:

- Personal data:
- Informed consent:

Dataset Result files:

- Personal data:
- Informed consent:

6. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

The data should be publicly available if possible and saved in a Findable, Accessible, Interoperable and Reusable (FAIR) manner, although this is not mandatory.