Support4LHS FAIR data questionnaire

The aim of this questionnaire is to create a catalogue of the datasets that will be produced as part of the Support4LHS project, in order to assess the necessary adjustments to the FAIR data publication framework developed as part of Objective 3.1 (Design and implementation of a FAIR Data Management Plan). The datasets will be published, to the extent possible, respecting the FAIR principles (Findable, Accessible, Interoperable, Reusable). The results of this questionnaire will be included in the Deliverable 3.1 (Specification of the Data Management Plan (DMP)) as a catalogue of datasets.

It is assumed that the leader of each Objective is aware of the datasets that the Objective will produce, so this form is aimed at Objective leaders. Even though it is impossible for an Objective leader to foresee all the datasets, the more details are provided herein, the less adjustments necessary for the FAIR data publication framework.

If your Objective will produce more than one dataset, fill the form for each dataset (But please try to group the data in as few datasets as possible).

Any questions can be directed to mikel.egana@ehu.eus.

* Derrigorrezkoa

Dataset ownership

Questions about who owns the data

1. What objective of the project are you the leader of? *

   Markatu aukera bakarra.

   ⬭ O1.1 Knowledge and standards-based transformation methods for process mining data ingestion in healthcare

   ⬭ O1.2 Extraction of knowledge graphs from clinical process models

   ⬭ O1.3 Interoperability framework based on knowledge graphs for clinical process models

   ⬭ O1.4 Knowledge graph methods for analysing clinical process models

   ⬭ O1.5 Exploitation of knowledge models for the explainability of process mining models

   ⬭ O2.1 New process mining methods and metaphors for the discovery of clinical processes from EHR data

   ⬭ O2.2 Integration framework for process mining models and clinical guideline models

   ⬭ O2.3 Interactive process mining methods based on integrated process mining models and clinical guideline models

   ⬭ O2.4 Analysis of compliance based on integrated process mining models and clinical guideline models

   ⬭ O3.2 Interactive process mining for the discovery of clinical processes in selected scenarios

   ⬭ O3.3 Development of knowledge models for selected scenarios

   ⬭ O3.4 Exploitation of process mining models and knowledge models in selected scenarios

2. Name, institution, and email of the person responsible (Objective leader) *

   _____

   _____

   _____

   _____

   _____

Dataset description

Basic information about the dataset

3. Name of the dataset *

4. Origin of the dataset *

5. Dataset purpose: what is the aim of this dataset? What is it going to be used   *
    for?

6. Dataset format: what is the expected format of the dataset? (File formats,   *
    relational databases, etc)

7. Data "shape": What are the data types of the dataset? (e.g. data types of the columns on a spreadsheet) *

_____

_____

_____

_____

_____

8. Data volume: what is the expected size of the dataset? *

_____

_____

_____

_____

_____

9. Updates: what is the expected update frecuency for the dataset? *

_____

_____

_____

_____

_____

Dataset interoperability

Information about the current interoperability level of the dataset

10. Is the dataset already published following FAIR principles? Where? *

_____

_____

_____

_____

11. Does the dataset reuse other datasets? Which ones? *

_____

_____

_____

_____

_____

12. Does the dataset include rich metadata? In which form? *

_____

_____

_____

_____

_____

13. Does the dataset follow naming conventions or any other community *
standards of its domain? Which ones?

_____

_____

_____

_____

_____

14. Does the dataset include vocabularies, ontologies, or any other standards at *
metadata or data level? Which ones?

_____

_____

_____

_____

_____

15. Does the dataset include new ontologies developed specifically for it? Where *
    are they published?

    _____

    _____

    _____

    _____

    _____

16. Is the dataset properly versioned? *

    _____

    _____

    _____

    _____

    _____

17. Does the dataset include an explicit and machine readable license? *

    _____

    _____

    _____

    _____

    _____

18. Are there any special methods or software needed to access the data? Which *
    ones?

    _____

    _____

    _____

    _____

    _____

Dataset
publication

Information about the specifics of publishing the dataset following FAIR
principles

19. Is there an embargo period in which the dataset, or its metadata, cannot be *
published? (For example if you prefer to wait for manuscript acceptance in a
journal)

_____

_____

_____

_____

_____

20. Do you agree to publish the dataset and/or its metadata following FAIR *
principles? Specify any limitations to publish the datasets and associated
metadata, detailing any alternative access methods. For example, in a clinical
dataset it might be impossible to publish the data to respect patients privacy,
but the method to access it is still valuable (For example the email of the
person in charge of the commitee to access the data might be provided as
basic metadata)

_____

_____

_____

_____

_____

21. Is the dataset protected by any regulation or special legislation related to *
personal data? Which one?

_____

_____

_____

_____

_____

22. Is the dataset going to be published in a specialised public repository (e.g. like   *
    UniProt –https://www.uniprot.org/–)? Which one? When is the expected
    publication date?

    _____

    _____

    _____

    _____

    _____

23. Has any anonymisation been applied to the dataset to preserve patient privacy?

    _____

24. I commit to post the dataset (or some of its metadata, in case of private data)   *
    and its updates to the shared resource that will be set up to that end (FTP
    server, OneDrive shared folder, etc.). (Name and email)

    _____