The NInFEA dataset is accessible from the Physionet repository: <https://doi.org/10.13026/c4n5-3b04>. This dataset has been presented as part of a peer-reviewed, scientific article by Pani, D et al. [Scientific Data](https://www.nature.com/sdata) 8:30 (2021) <https://doi.org/10.1038/s41597-021-00811-3> which is entitled "NInFEA: Non-Invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research". The abstract is given below:

Non-invasive foetal electrocardiography (fECG) continues to be an open topic for research. The development of standard algorithms for the extraction of the fECG from the maternal electrophysiological interference is limited by the lack of publicly available reference datasets that could be used to benchmark different algorithms while providing a ground truth for foetal heart activity when an invasive scalp lead is unavailable. In this work, we present the Non-Invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research (NInFEA), the first open-access multimodal early-pregnancy dataset in the field that features simultaneous non-invasive electrophysiological recordings and foetal pulsed-wave Doppler (PWD). The dataset is mainly conceived for researchers working on fECG signal processing algorithms. The dataset includes 60 entries from 39 pregnant women, between the 21st and 27th week of gestation. Each dataset entry comprises 27 electrophysiological channels (2048 Hz, 22 bits), a maternal respiration signal, synchronised foetal trans-abdominal PWD and clinical annotations provided by expert clinicians during signal acquisition. MATLAB snippets for data processing are also provided.

Crucially for this guide, the NinFEA publication includes the details of methodology to explain how the data set was created and how it can be used for calibration. It also provides machine-accessible metadata for the reported data set at the study level: <https://doi.org/10.6084/m9.figshare.13283492>. This study level metadata is represented in CSV (human) and JSON (machine) using numerous sources of terminology.

The scientific publication, the data set, and the associated metadata were reviewed against the RDA criteria for FAIR assessment (see earlier section). The overall assessment at the study level, is that arguably all essential FAIR assessment criteria are met, which is no surprise given the fact that the dataset is provided as part of Physionet, which is supported by a strong international community. However, we have identified areas for improvement especially regarding interoperability which will benefit from application of the HL7 FHIR standards.

Furthermore, the HL7 FHIR standards enable more detailed representation of FAIR metadata at the anonymised patient level for method and context which is currently limited to the full text of the publication (HTML and PDF formats), which is not readily readable by machine. Making data FAIR is an iterative process guided by the priority of the RDA maturity indicators. Optimal level of FAIRness should be determined by understanding the requirements of the community (rather than perfection!). Here we we are going to show what is possible using the current HL7 FHIR standards, how these can bring unique and greatest value and how HL7 FHIR-based solutions compare to alternative technical options for FAIR Implementation.

### NinFEA: HL7 FHIR enabled FAIR improvements

Taking a curator approach, the features that the utilisation of the HL7 FHIR standards would improve FAIR implementation have been analysed, even further for this NInFEA dataset. As well as how metadata could be enhanced in a FAIR manner at the study and patient levels.

Observations made during the initial FAIR assessment:

1. The metadata provided for the data set can be further extended, by harnessing the full text of the open-source paper (<https://doi.org/10.1038/s41597-021-00811-3>).
2. Richer metadata would facilitate future reuse of the dataset to maximise it's value which is the major motivation for making data FAIR.

Recommendations for FAIR data improvement enabled by HL7 FHIR:

1. Enriching the metadata using the HL7 FHIR standards is likely to bring unique and greatest value. HL7 FHIR enhancement to make the metadata more FAIR could be scaled through automation.
2. Developing HL7 FHIR resources and storing the data set in a HL7 FHIR server would allow easier access to more researchers, further pushing the envelope for interoperability, going beyond Physionet.
3. Information on data provenance and license would be included in the metadata with a machine-readable header on the HL7 FHIR server.
4. Publishing the structure and elements as FAIRified data and metadata provides richer support for the work done, thereby building best practices.

Overall, the purpose of application for the dataset should drive finding the optimal balance between acceptable cost and richness of metadata. We have started this from the study level, provided already by Physionet and enriched further from the full text paper, which is an open-source publication. We will explore the library HL7 FHIR resource ([https://www.hl7.org/HL7 FHIR/library.html](https://www.hl7.org/fhir/library.html)) to follow-up on the initial FAIR assessment, as noted in the list of observations. Other technical options using the HL7 FHIR standards will be considered and explored as well.