### NInFEA: Non-Invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research

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 FHIR standards. Furthermore, the 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 interactive 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 are going to show what is possible using the current FHIR standards, how these can bring unique and greatest value and how FHIR-based solutions compare to alternative technical options for FAIR Implementation.

### NinFEA: FHIR enabled FAIR improvements (work in progress)

Taking a curator approach, we looked more closely at the features that would push FAIR implementation even further for this NInFEA dataset. Notably, significant effort has been invested on preparing the data set.  What FAIR improvements could be made by implementation of the FHIR standards? Can the FAIR metadata be enhanced at the study and patient levels?

Several observations were made during the initial FAIR assessment:

1. The metadata provided for the data set can be further extended, especially by harnessing the open source, full text paper.
2. Richer metadata would facilitate repurposing for future reuse of the dataset.
3. What is possible to enrich the metadata using FHIR? What brings unique and greatest value? Approaches to automate FHIR enhancement if FAIR metadata?
4. Developing FHIR resources and storing the data set in a FHIR server would allow easier access to more researchers, further pushing the envelope for interoperability, going beyond Physionet.
5. Information on data provenance and licence could be included in the metadata with a clear header rather than in comment.
6. Publishing information on the structure of the data and the clinical information further support the work done, building best practices.

Overall, there is currently a tradeoff between the quantity and richness of metadata at the study level, provided for the dataset on Physionet. We will explore the library FHIR resource (<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 will be considered and explored as well.