# Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** Feasibility of machine learning for seismic forward modelling and inversions

**Creator:**Abdoulie Faal

**Principal Investigator:** Ivan Lokmer

**Data Manager:** Conall Evans

**Affiliation:** University College Dublin

**Template:** Health Research Board DMP Template

**Project abstract:**

Imaging volcanic interiors is of paramount importance for understanding volcano-seismic signals
and their underlying sources. However, determining fine scale structure in highly heterogeneous
media is a significant challenge using traditional imaging approaches. Furthermore, modelling and
inversion tools often employ cumbersome and lengthy procedures, which can be slow to implement, especially during volcanic crises when results are needed swiftly as large data volumes arrive at Volcano Observatories. Machine-learning (ML) methods, which have experienced rapid growth over the last decade, have strong potential to address this challenge
due to their suitability for complementing physics-based numerical simulations and inversion. In
particular, we examine the feasibility of imaging small-scale heterogeneities beneath volcanoes,
such as propagating individual dykes, directly from seismic data using rapid ML-based imaging.

Here we build on previous work where a large suite (> 5000) of seismic earthquake gathers (i.e.
seismic records from individual earthquakes) derived from numerical simulations in highly heterogeneous 2D velocity models, were used to train a Fourier Neural Operator (FNO). Subsequently that FNO was used to invert for complex structure in previously unseen geologically
realistic 2D models. As the training procedure is extremely computationally expensive, and is likely
prohibitive in 3D, here we ask: “can meaningful information be retrieved from seismic data derived
from 3D simulations, based on an FNO that was trained only on 2D seismic data”? We see the answer to this question as important, as it helps determine the nature of the FNO training required in order to apply this new methodology beyond the numerical domain into the 3D physical world.

**ID:** 195191

**Start date:** 26-09-2025

**End date:** 26-09-2027

**Last modified:** 19-01-2026

**Grant number / URL:** iCRAG2

**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

# Feasibility of machine learning for seismic forward modelling and inversions

## Data description and collection or re-use of existing data

### How will new data be collected or produced and/or how will existing data be re-used?

Data will be generated in python and using SPECFEM3D

### What data (for example the kind, formats, and volumes), will be collected or produced?

The data will be produced digitally and will be synthetic
Formats:
.py files
.dox files
.dat files
.npy files
.mseed files
.png files

## Documentation and data quality

### What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

A README notebook file can accompany the data to provide some guidance

### What data quality control measures will be used?

It will be generated by me entirely and will contain no personal information

## Storage and backup during the research process

### How will data and metadata be stored and backed up during the research process?

All the data will be periodically backed up to google drive

**How will data security and protection of sensitive data be taken care of during the research?**

There will be no sensitive data

## Legal and ethical requirements, codes of conduct

**If personal data are processed, how will compliance with legislation on personal data and on security be ensured?**

None

**How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**

None

**What ethical issues and codes of conduct are there, and how will they be taken into account?**

None

## Data sharing and long-term preservation

**How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**

No restrictions. Data will be shared upon request

**How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**

Final data sets may be stored for long term preservation in a data repository.

**What methods or software tools are needed to access and use data?**

specfem3d, python or matlab

**How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Any data stored for long term preservation will be handled on consultation with the iCRAG data manager.

## Data management responsibilities and resources

**Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

Conall Evans will be responsible

**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)?**

Near the end of the project, I will dedicate some time to ensuring the data is structured and preserved in a useful and accessible manner.