



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **dispel4py: An Open Source Python Framework for Encoding, Mapping and Reusing Seismic Continuous Data Streams: Intensive Analysis and Data Mining**

**Citation for published version:**

Filgueira Vicente, R, Krause, A, Atkinson, M, Spinuso, A, Klampanos, I, Magnoni, F, Casarotti, E & Vilotte, J-P 2015, dispel4py: An Open Source Python Framework for Encoding, Mapping and Reusing Seismic Continuous Data Streams: Intensive Analysis and Data Mining. in American Geoscience Union Fall Meeting (AGU 2015).

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

American Geoscience Union Fall Meeting (AGU 2015)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## dispel4py : An Open Source Python Framework for Encoding, Mapping and Reusing Seismic Continuous Data Streams: Intensive Analysis and Data Mining.

Scientific workflows are needed by many scientific communities, such as seismology, as they enable easy composition and execution of applications, enabling scientists to focus on their research without being distracted by arranging computation and data management. However, there are challenges to be addressed. In many systems users have to adapt their codes and data movement as they change from one HPC-architecture to another. They still need to be aware of the computing architectures available for achieving the best application performance.

We present dispel4py, an open-source framework presented as a Python library for encoding and automating data-intensive scientific methods as a graph of operations coupled together by data-streams. It enables scientists to develop and experiment with their own data-intensive applications using their familiar work environment. These are then automatically mapped to a variety of HPC-architectures, i.e., MPI, multiprocessing, Storm and Spark frameworks.

dispel4py comes with data provenance, and with an information registry that can be accessed transparently from within workflows. dispel4py has been enhanced with a new run-time adaptive compression strategy to reduce the data stream volume and a diagnostic tool which monitors workflow performance and computes the most efficient parallelisation to use.

dispel4py has been used by seismologists in the project VERCE for seismic ambient noise cross-correlation applications and for orchestrated HPC wave simulation and data misfit analysis workflows; two data-intensive problems that are common in today's research practice. Both have been tested in several local computing resources and later submitted to a variety of European PRACE HPC-architectures (e.g. SuperMUC & CINECA) for longer runs without change. Results show that dispel4py is an easy tool for developing, sharing and reusing data-intensive scientific methods.