



## **Provenance tracking for scientific software toolchains through on-demand release and archiving**

David Ham

Imperial College London, Department of Mathematics, London, United Kingdom (david.ham@imperial.ac.uk)

There is an emerging consensus that published computational science results must be backed by a provenance chain tying results to the exact versions of input data and the code which generated them. There is also now an impressive range of web services devoted to revision control of software, and the archiving in citeable form of both software and input data.

However, much scientific software itself builds on libraries and toolkits, and these themselves have dependencies. Further, it is common for cutting edge research to depend on the latest version of software in online repositories, rather than the official release version. This creates a situation in which an author who wishes to follow best practice in recording the provenance chain of their results must archive and cite unreleased versions of a series of dependencies.

Here, we present an alternative which toolkit authors can easily implement to provide a semi-automatic mechanism for creating and archiving custom software releases of the precise version of a package used in a particular simulation. This approach leverages the excellent services provided by GitHub and Zenodo to generate a connected set of citeable DOIs for the archived software. We present the integration of this workflow into the Firedrake automated finite element framework as a practical example of this approach in use on a complex geoscientific tool chain in practical use.