

Direct DOI Dataset Access

Associating datasets with persistent identifiers (a globally unique ID) is crucial for ensuring that research outputs are available for open science and comply with the FAIR principles (Findable, Accessible, Interoperable, and Reusable). This enhances the value of the data and it also allows other people to recreate existing work, or to use the data in new and novel research.

Currently, several platforms and institutes offer the possibility to store data with a DOI, a type of persistent identifier. However, no standard exists that allows a machine to access or download the data from a DOI. Repositories typically either have no way to download data automatically, or have adopted some proprietary solution. With no widely deployed standard, any support for accessing or downloading data from a DOI can only be incomplete.

You would be working on this task, for which a solution has already been proposed, based on existing approaches in other contexts (HTTP content-negotiation, and a dataset description language). You would implement a proof-of-concept code, based on the design, that will demonstrate that the approach can work. This will involve a mixture of modifying existing production software and developing new code to build a demonstration of the benefits of this approach.

Group

IT

Project Category

B5. Computing

DESY Site

Hamburg

Special Qualifications

Requirements: Experience with programming and git

Beneficial: network protocols and HTTP in particular, knowledge of golang, experience with filesystems and FUSE.

Primary authors: MILLAR, Alexander Paul (IT (Research and Innovation in Scientific Co)); Dr SCHWARZ, Kilian (IT (IT Scientific Computing)); NENTWICH, Melanie (None)

Presenter: MILLAR, Alexander Paul (IT (Research and Innovation in Scientific Co))