ES-doc-errata: an issue tracker platform for CMIP6

Atef Ben Nasser, Guillaume Levavasseur, Mark Greenslade, and Sébastien Denvil
IPSL, Paris, France

In the context of overseeing the quality of data, and as a result of the inherent complexity of projects such as CMIP5/6, it is a mandatory task to keep track of the status of datasets and the version evolution they sustain in their life-cycle.

The ESdoc-errata project aims to keep track of the issues affecting specific versions of datasets/files. It enables users to resolve the history tree of each dataset/file enabling a better choice of the data used in their work based on the data status.

The ES-doc-errata project has been designed and built on top of the Parent-IDentifiers handle service that will be deployed in the next iteration of the CMIP project, by ensuring maximum usability of ESGF ecosystem and encapsulated in the ES-doc structure. Consuming PIDs from handle service is guided by a specifically built algorithm that extracts meta-data regarding the issues that may or may not affect the quality of datasets/files and cause newer version to be published replacing older deprecated versions. This algorithm is able to deduce the nature of the flaws to the file granularity, that is of high value to the end-user.

This new platform has been designed keeping in mind usability by end-users specialized in the data publishing process or other scientists requiring feedback on reliability of data required for their work. To this end, a specific set of rules and a code of conduct has been defined. A validation process ensures the quality of this newly introduced errata meta-data , an authentication safe-guard was implemented to prevent tampering with the archived data, and a wide variety of tools were put at users disposal to interact safely with the platform including a command-line client and a dedicated front-end.