IS-ENES3 Deliverable D7.5

Report on operational support for CMIP documentation

Reporting period: 01/01/2022 – 31/03/2023

Authors: G. Levavasseur (CNRS-IPSL), D. Hassell (UREAD-NCAS)

Reviewers: U. Fladrich (SMHI), S. Joussaume (CNRS-IPSL)

Release date: 05/01/2023

ABSTRACT

This deliverable reports on the ES-DOC operation support to CMIP. We describe the operational ES-DOC eco-system deployed as well as all support actions put in place to produce and maintain the CMIP documentation. This document also describes lessons learnt from CMIP6, the remaining support actions and recommendations to follow in order to maintain the ES-DOC service after the end of the IS-ENES project

Dissemination Level		
PU	Public	X
CO	Confidential, only for the partners of the IS-ENES3 project	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Revision Table			
Version	**Date**	**Name**	**Comments**
Document Structure and Contributors	2022-06-20	Guillaume Levavasseur	Preliminary Structure
0.1	2022-11-22	Guillaume Levavasseur	Additional content
0.2	2022-11-30	David Hassell	Comments
0.3	2022-12-16	Sylvie Joussaume	External review
0.4	2022-12-19	Uwe Fladrich	External review
1.0	2022-12-19	Guillaume Levavasseur	Comments from review addressed.

Table of contents

List Of Images

List Of Tables

Executive Summary

Since the fifth phase of the Coupled Modeling Intercomparison Project (CMIP), the WCRP Working Group on Coupled Modelling (WGCM) and its Infrastructure Panel (WIP) have tasked the Earth System Documentation (ES-DOC) project with documenting all aspects of the sixth phase of CMIP.

The ES-DOC software ecosystem facilitates both the provision and the consumption of documentation of the CMIP6 workflow and, where possible, automates the various and often complex stages involved. It offers services for metadata search, comparison and creation, following the CIM standard (Common Information Model[1]) and provides an environment to document the modeling workflow.

ES-DOC brands two main services, which are (i) a Documentation Service for CMIP models, simulations and protocols and (ii) an Errata Service for centralizing known issues. In 2018, both ES-DOC services became operational for CMIP6. Through the IS-ENES3 project, the ES-DOC services are hosted on remote virtual machines provided by OpalStack[2].

This document provides a review of the ES-DOC operational support put in place to produce and maintain the CMIP documentation. The document addresses the deliverable D7.5 "Report on operational support for CMIP documentation" of the IS-ENES3 project, within the WP7/VA2 "Data standards, distribution and processing services".

The work presented in the report is packaged in different sections, which describe first briefly the ES-DOC ecosystem and architecture. We then report on the different support actions to maintain the operational services. This may concern the machines themself or the documentation produced by the climate modeling groups to be ingested in the ES-DOC system. Finally, we describe the user support actions run to accompany the climate modeling groups in the documentation production and handling of user requests.

1. Introduction

The ES-DOC project aims to nurture an ecosystem of tools and services in support of Earth System Model documentation. The purpose of this suite of tools is to enable the scientific community to better understand and utilize Earth System Model (ESM) data by describing all aspects of CMIP6 through a set of documents illustrated in Figure 1.

[1] https://en.wikipedia.org/wiki/Common_Information_Model_(computing)
[2] https://www.opalstack.com/

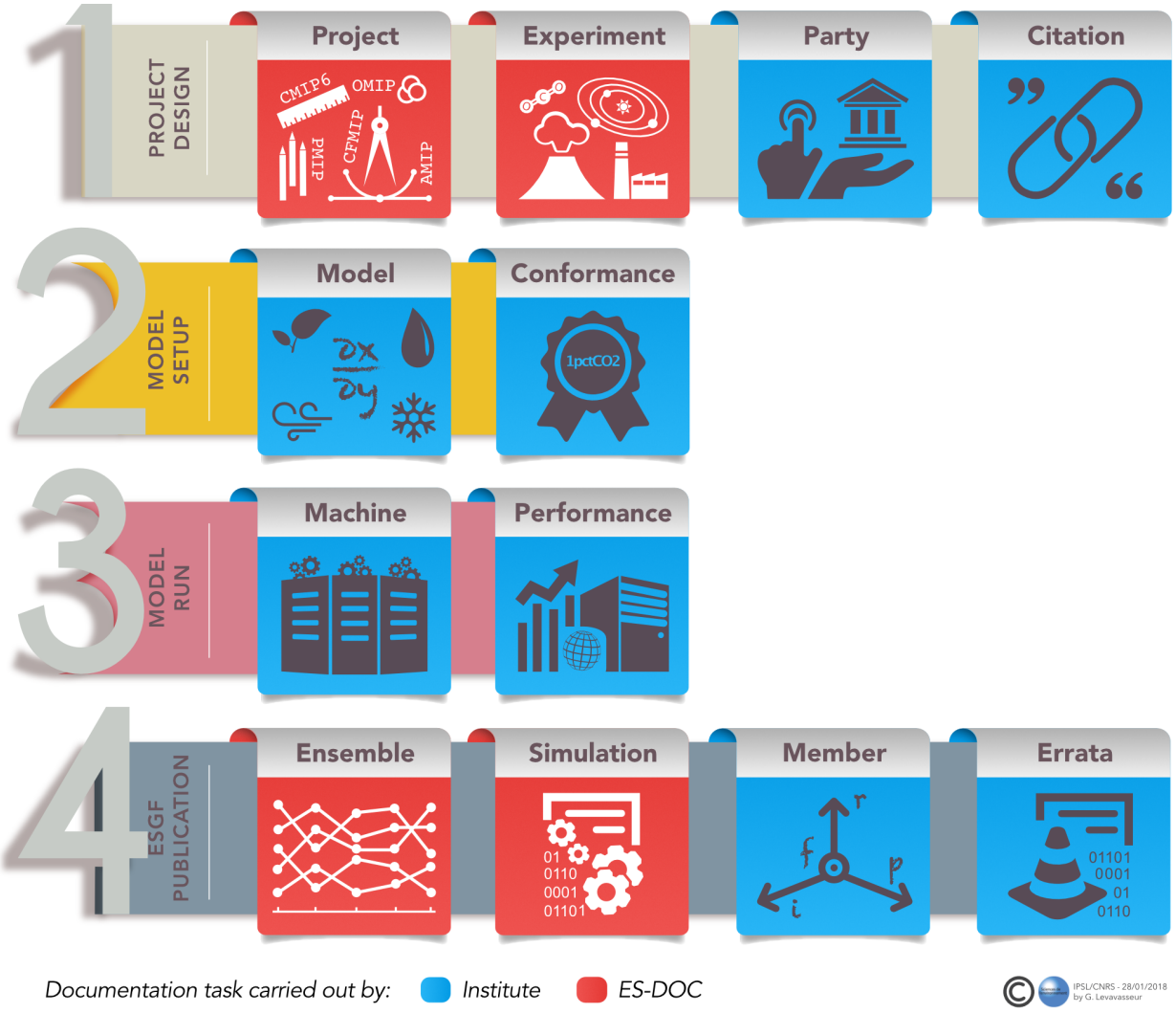

Figure 1: ES-DOC workflow and CIM documents

Building on CIM (Common Information Model) concepts[3], the Metafor project developed the CIM1 formalism[4]. The documents (in bold) in Figure 1 are created for the CMIP6 **Project** based on the CIM2 developed during the IS-ENES2 project. A **Simulation** is a run of a configured **Model** which has **Conformance** to the numerical requirements of an **Experiment**, runs on a **Machine** with an HPC **Performance** and produces output Datasets (archived on the Earth System Grid Federation - ESGF). An **Ensemble** is a set of Simulations whose Axis **Members** describe how the simulations differ. A **Party** describes a person or organization involved in the modeling process

[3] https://fr.wikipedia.org/wiki/Common_Information_Model
[4] Describing Earth system simulations with the Metafor CIM, Lawrence et al. 2012, 10.5194/gmd-5-1493-2012

and a published reference is recorded in a **Citation**. The IS-ENES 3 project funded support to get those CIM2 documents from CMIP6 climate modeling group.

ES-DOC coordinates with community efforts such as the CMIP Infrastructure Panel (WIP), the ESGF and more recently with the CMIP International Project Office. Thus, the ES-DOC **Errata** Service has been the community's answer to an issue raised by the complexity of projects like CMIP5 and CMIP6, namely to provide a platform that enables users to record and track reasons that motivate dataset version changes. Datasets may be associated with **Errata** if any problems are discovered after publication to ESGF.

2. Equipment & brief reminder of the system architecture

The ES-DOC web-services and applications are deployed on servers provided by a third party company OpalStack. Two servers have been provisioned with an almost identical setup. Whilst server one acts as a TEST environment, server two acts as a PROD environment.

The services being provided by the ES-DOC software stack are depicted in the workflow diagram of Figure 38 of report D10.1[5] and are detailed in report D10.2[6]. In summary, these are:

- A website hosted with WordPress (https://es-doc.org/), supported by dedicated servers and databases on the OpalStack commercial cloud hosting.
- A software ecosystem and archive contained in GitHub repositories under the 'ES-DOC' and 'ES-DOC-INSTITUTIONAL' organizations enable content pushed by modeling institutes to be processed and made available on the WordPress website.
- Python-based utility libraries to automate the creation and publication of standardized documents and controlled vocabularies, and the storage of documents in repositories on GitHub.
- A shell-script library to facilitate development and maintenance.
- Python web services to manage documentation and errata stored in the Opalstack databases.

[5] Fiore, S., Nassisi, P., Antonio, F., Barring, L., Ben Nasser, A., Berger, K., Hassell, D., Juckes, M., Kershaw, P., Kindermann, S., Levavasseur, G., Nuzzo, A., Pagé, C., Stephens, A., Som de Cerff, W., Spinuso, A., Stockhause, M., & Weigel, T. (2020). Architectural document of the ENES CDI software stack (D10.1). Zenodo. https://doi.org/10.5281/zenodo.4309892

[6] Spinuso, Alessandro, Som de Cerff, Wim, Nassisi, Paola, & Pagé, Christian. (2021). First release of the ENES CDI software stack (D10.2). Zenodo. https://doi.org/10.5281/zenodo.4450012

- Web applications written in JavaScript and as Vue.js components that support the viewing, searching, and comparing of the published documentation, as well as serving and displaying other relevant content:
    - Documentation Service: https://search.es-doc.org/
    - Errata Service: https://errata.es-doc.org/
    - Dedicated documentation explorer: https://explore.es-doc.org/ that renders CIM 2[7] and CIM1 documents and CMIP6 further URL information[8].

ES-DOC services are fully interoperable with the ESGF stack: UREAD-NCAS develops the Python library "cdf2cim" included into the ESGF publisher; the CNRS-IPSL is part of the CMIP Data Node Operation Team that orchestrate ESGF developments and deployments.

All elements of this generic workflow have been implemented and are fully available as services, software packages and specification documents available from the ES-DOC organization GitHub repositories at https://github.com/ES-DOC. The table 1 below details the description of aforementioned repositories.

Component	Description	Documentation	GitHub repository	Version
Authorization management	Suite of GitHub repositories managing authorisation role for ES-DOC content.	-	https://github.com/ES-DOC-INSTITUTIONAL	-
CIM schema	CIM document for Earth system documentation model	https://technical.es-doc.org/	https://github.com/ES-DOC/esdoc-cim-v2-schema	2.2
pyesdoc	Python client for ES-DOC	https://technical.es-doc.org/	https://github.com/ES-DOC/esdoc-py-client	0.14.2.0
pyessv	Python library to manager controlled vocabulary for ES-DOC	https://technical.es-doc.org/	https://github.com/ES-DOC/pyessv	0.8.4.3
pyessv-archive	Centralized controlled vocabulary archive for ES-DOC	https://technical.es-doc.org/	https://github.com/ES-DOC/pyessv-archive	master
pyessv-ws			https://github.com/ES-DOC/pyessv-ws	
pyessv-config			https://github.com/ES-DOC/pyessv-config	
cdf2cim	Python library to publish simulation CIM documents for ESGF published data.	https://technical.es-doc.org/	https://github.com/ES-DOC/esdoc-cdf2cim	1.0.3.0

[7] https://github.com/ES-DOC/esdoc-web-explorer/tree/master/src/components/cim2
[8] https://github.com/ES-DOC/esdoc-web-explorer/tree/master/src/components/cmip6/further-info

-specializations-			https://github.com/ES-DOC/*-specializations-*	
infrastructure	Documentation on the technical underpinning of the ES-DOC project		https://github.com/ES-DOC/esdoc-infrastructure	
machine-fe	ES-DOC machine front end		https://github.com/ES-DOC/esdoc-machine-fe	
shell	ES-DOC command line shell for deployment		https://github.com/ES-DOC/esdoc-shell	
web-compare			https://github.com/ES-DOC/esdoc-web-compare	
web-search			https://github.com/ES-DOC/esdoc-web-search	
web-explorer			https://github.com/ES-DOC/esdoc-web-explorer	
web-view			https://github.com/ES-DOC/esdoc-web-view	
web-view-specialization			https://github.com/ES-DOC/esdoc-web-view-specialization	
documentation-ws				
errata-ws	Errata Web-service	https://technical.es-doc.org/	https://github.com/ES-DOC/esdoc-errata-ws	master
errata-fe	Errata front-end and forms	https://technical.es-doc.org/	https://github.com/ES-DOC/esdoc-errata-fe	master
errata-client	Errata CLI to manage issue life-cycle	https://es-doc.github.io/esdoc-errata-client/	https://github.com/ES-DOC/esdoc-errata-client	2.3.1

Table 1: List of repositories of ES-DOC components

In addition, technical documentation is now available at https://technical.es-doc.org/. It describes all technical aspects of the ES-DOC stack from the libraries developed to the Command-Line-Interface applications.

Finally, all the installation procedures have been finalized and documented through the repository 'DEVOPS' (https://github.com/ES-DOC/devops). It allows the ES-DOC eco-system to be deployed on other machines if necessary as it occurred in the past when ES-DOC services migrated from WebFaction to OpalStack provider.

3. Operational support

3.1. Service level targets & KPIs

This section focuses on the collection of key performance indicators (KPIs) for the ES-DOC services within the ENES climate data infrastructure and their evolution since the start of IS-ENES3. All KPIs are summarized in the Table 2 below.

ES-DOC Service	KPI	RP1 value	RP2 value
Documentation Service	*Number of institution liaised*	49	83
	Number model documented	22	36
	Number of visits on ES-DOC Search & Explorer	21,405	NA
	Number of CIM simulation documents generated	2,462,979	2,813,991
Errata Service	*Number of ESGF affected datasets*	89,993	244,623
	Number of registered issues	198	378
	Number of resolved issues (including WontFix status)	107	272
	Number of visits on Errata front-end	5,705	15,076

Table 2: List and evolution of ES-DOC KPIs

3.2. Service requests and incident handling

OpalStack IT services according to services described in Section 2 are delivered during 24 hours per day, 7 days per week (i.e., 365 days or 8,760 hours). OpalStack is responsible for machine redundancy and local storage management depending on the subscription plan. Planned and announced interruptions may reduce the effective operating time of a service, nevertheless some unexpected incidents on the OpalStack side led to a disruption of the ES-DOC services.

Therefore, CNRS-IPSL and UREAD-NCAS are responsible for system monitoring and failover in a best effort fashion. Incidents are dealt with as far as possible within 48 hours , regardless of the

impact of the outage or service quality degradation. If an incident is raised by a user through one of the communication channels (see section 4), the user receives a quick acknowledgement and then we communicate on the user mailing list, whenever possible, an early evaluation of the impact.

Incidents are then handled as quickly as possible (ideally within a maximum of 3 days) depending on the workload of UREAD-NCAS and CNRS-IPSL staff and the impact of the outage or service quality degradation.

4. Liaison and escalation

The ES-DOC services are historically designed for two kind of users:

- The **modeling centers** that **need to provide** their model documentation and information about known data issues (see number of modeling groups liaised in Table 2).
- The **data users** that **need to know** about model documentation and information about data issues (see numbers of visits on ES-DOC websites in Table 2).

Thus, ES-DOC services have two support objectives:

- Accompanying the modeling groups in their documentation process in a form of long-term liaison.
- Guiding the data users in ES-DOC documentation through a dedicated mailing list and user documentation.

4.1. Liaison with modeling groups

In 2018, during the ESGF Data Challenge that preceded the publication of the CMIP6 data, the ES-DOC team discussed with all the modeling groups to identify their ES-DOC officer(s) and describe to them the publication workflow[9] which involves first documenting the model whose data is to be published.

Major efforts have been made between 2019 and 2021 to encourage and support CORDEX institutions in the same way to document their Regional Climate Models (RCMs) and known issues within the ES-DOC ecosystem.

[9] https://pcmdi.llnl.gov/CMIP6/Guide/modelers.html

ES-DOC relies on a strong collaboration and liaison with the CMIP and CORDEX data providers that spend time to fill the documentation about their models and annotate any known issues about their datasets. Liaison with data providers is ensured in the same way by UREAD-NCAS for the CMIP6 modeling workflow documentation (from experiments to models to simulations) and CNRS-IPSL for errata information. Both services rely on the same user management delegate to the GitHub OAuth through the "ES-DOC-INSTITUTIONAL" GitHub organization.

4.2. Data user support

The modeling groups are supported via the ES-DOC liaison at each CMIP6 modeling institute with identified creators. Identified creators (also called "ES-DOC officers") have been trained by ES-DOC staff and organize the creation of documentation locally. Their primary contact with ES-DOC is via the support email (support@es-doc.org) and a dedicated ES-DOC-liaison mailing list.

The same support email (support@es-doc.org) is used for data user support. During IS-NES3 projet, data user requests were mainly about:

- Guidance through the web interfaces of the ES-DOC services,
- Missing or unavailable model documentation,
- Reporting of data issues.

ES-DOC support is provided during 5 days per week without on-call service. Any message sent to the ES-DOC mailing list will be forwarded to all ES-DOC staff (from UREAD-NCAS and IPSL-CNRS) through a dedicated Slack domain with specific hooks. Table 3 summarizes the metrics of ES-DOC user support.

ES-DOC Service	Indicators	RP1 value	RP2 value
Documentation Service	*Number of email received on support*	80	45
Errata Service	*Number of email received on support*	40	91

Table 3: ES-DOC user support metrics

5. Conclusion and recommendations

In the Open Science context, FAIR data must have a high level of documentation. The ES-DOC provides a necessary ecosystem of tools to centralize, standardize and compare documentation about CMIP6 data. This is a clearly achieved objective by ES-DOC that developed and deployed within IS-ENES a set of services allowing modeling groups and users to deal with different kinds of data documentation. Indeed, ES-DOC now offers a proper handling of errata information through a user-friendly search interface and dedicated API, as well as a comprehensive CIM document Search & Explorer for learning about CMIP6 MIPs, experiment and models.

However, the cost of accompanying the modeling groups to appropriate the ES-DOC tools has been underestimated. The documentation of CMIP6 models is incomplete and the documentation of known errors is not systematic, which reduces the interest of users in ES-DOC services because they do not find exhaustive information.

In 2021, an ESGF user survey included several questions about the use of ES-DOC services to collect all users' needs and requirements. Many users already suggested improvements to the services or new features to take into account lessons learnt from CMIP6. These new features have already been discussed together with ES-DOC, WIP and WGCM and will be developed and deployed during RP3.

For instance, the Errata Service has been completely redesigned and improved to allow each user to report problems through a data provider moderation and notification mechanism. This should reduce the stress on the latter to report and describe every issue encountered.

ESGF users also mentioned the need to simplify access to the documentation from the data itself, which is considered a bit painful with a lot of browsing and clicking. This requirement has been addressed by the implementation of the CMIP6 “further_info_url” that links each CMIP6 datasets to the corresponding ES-DOC model documentation. Although the errata pages are linked to ESGF via the use of Persistent IDentifiers (PIDs) assigned to the data, the access to corresponding issues could be tied to ESGF data services.

With the end of the IS-ENES3 project, the issue of continuity, maintenance of ES-DOC services and their scalability towards CMIP7 must be addressed. ES-DOC services will continue to be operated by a strong collaboration between UREAD-NCAS and CNRS-IPSL staff. Only the operational maintenance will be provided by in-kind contribution of both institutes. To this end, it is currently being discussed to migrate ES-DOC services to dedicated VMs at UREAD-NCAS and/or CNRS-IPSL site(s). A load-balancing approach (e.g., in the same way as deployed through Amazon Web Service for Copernicus data provision) could be put in place to ensure redundancy

of services in case of failure at one site. The documentation of the ES-DOC software stack itself has been brought forward and has been a considerable amount of work. The purpose of this documentation is to enable anyone to be able to maintain and develop the ES-DOC services in the future, if (or when) the current ES-DOC team is not able to support the project any longer. In addition, the knowledge transfer between existing staff will also help provide a better level of maintenance beyond the IS-ENES3 project. Finally, scalability towards CMIP7 relies mainly on the external funding that will be available to :

1. Improve the interfaces dedicated to the modeling groups and facilitate the creation of documentation, in particular that of the models.
2. Dedicate human resources to accompany the modeling groups in the documentation processes.