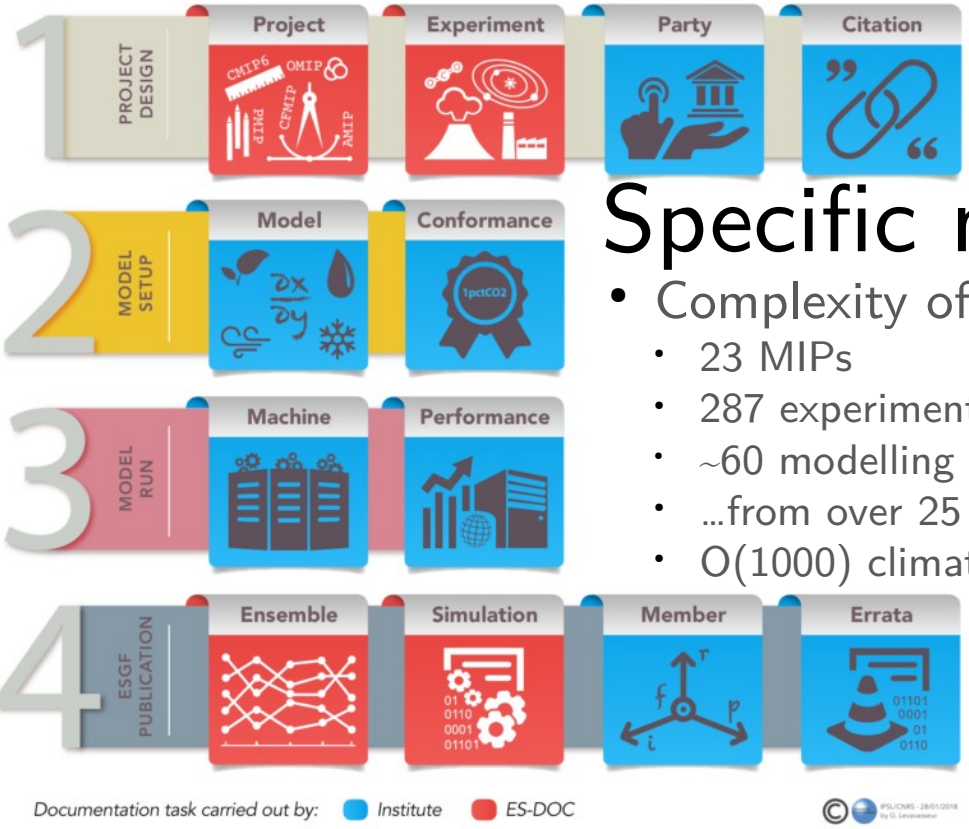
The ES-DOC website and Further Info URL as portals for exploring CMIP6

Sadie Bartholomew
University of Reading and National Centre for Atmospheric Science (NCAS)

On behalf of the wider ES-DOC team

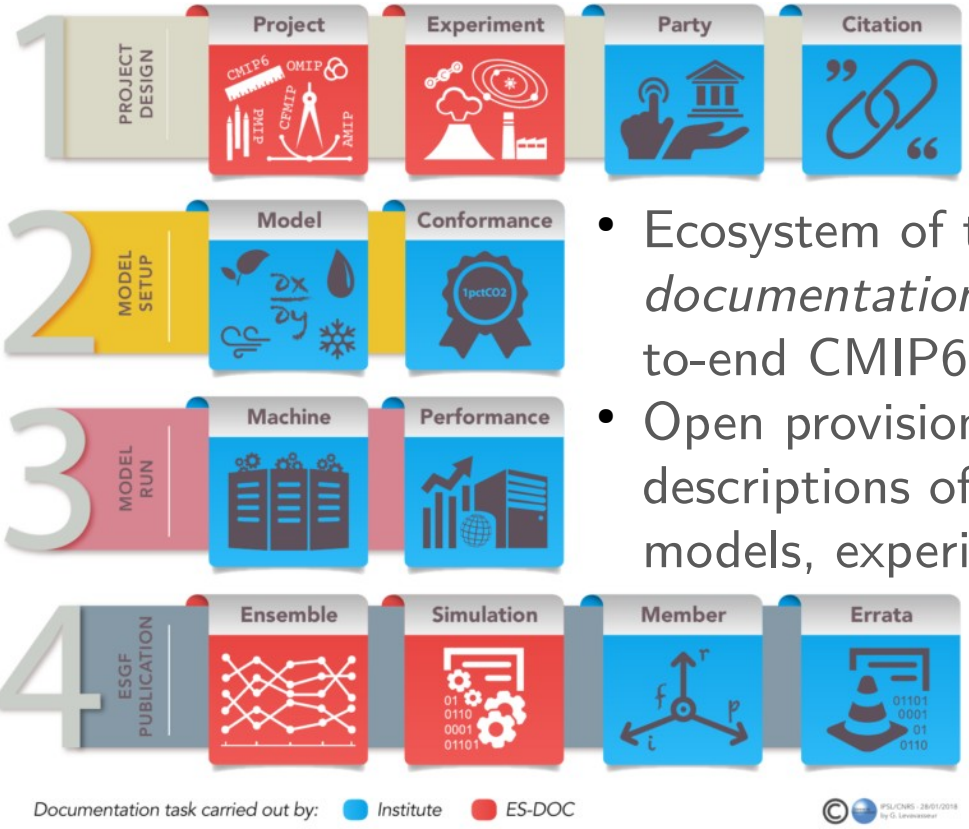
ES-DOC context

- ES-DOC: the *Earth System Documentation* project
- Global project going since ~2013 (IS-ENES-2 and -3 funds)
- Aims to enable the scientific community to better **understand** and **utilise** Earth system model data by developing and advancing tools and services in support of relevant documentation **collection**, **generation** and **access**
  - streamlines collection process, aiming to auto-generate where possible
- Reflects a growing understanding that large-scale dataset archives needs a high-quality documentation ecosystem
- Sprang from other projects with aligned aims:
  - METAFOR (European/EU-funded), 2008 - 2011
  - Earth System Curator (US-led), even earlier than Metafor



Specific motivation

- Complexity of CMIP6 workflow:
  - 23 MIPs
  - 287 experiments
  - ~60 modelling institutions ("parties")...
  - ...from over 25 countries
  - O(1000) climate variables



ES-DOC for CMIP6

- Ecosystem of tools to support *documentation* covering full end-to-end CMIP6 workflow
- Open provision of structured descriptions of the CMIP6 MIPs, models, experiments, simulations, processes and outputs.

es-doc
Earth System Documentation

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

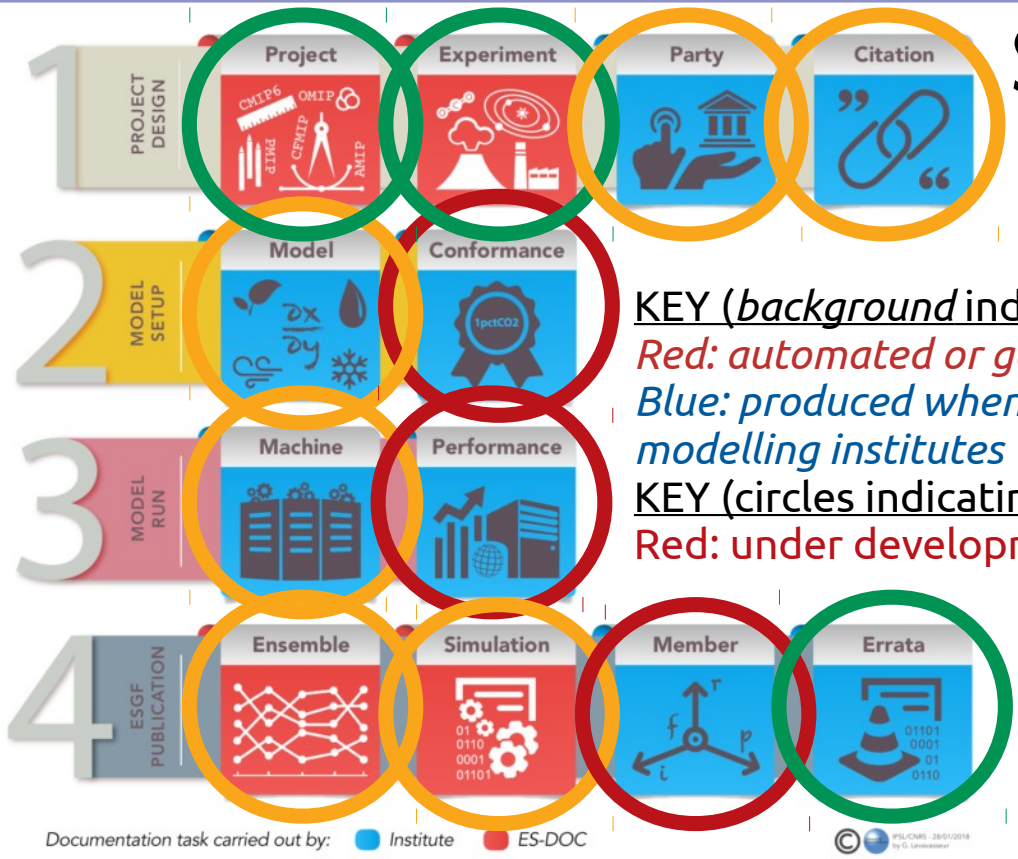
University of Reading

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Status for CMIP6

- ES-DOC has been operational for CMIP6 since Dec. 2018
  - published documentation is available, new additions constantly added
- Some CMIP6 components are fully documented, others awaiting input, when ready, from modelling groups
- All published CMIP6 documentation is readily available to view, explore and compare by any interested party
  - means to do this (efficiently) are covered later in this talk!

Status for CMIP6



KEY (*background* indicating task form)

Red: automated or generated by ES-DOC

Blue: produced whenever ready by modelling institutes

KEY (circles indicating status)

Red: under development, due later 2021

Yellow: ongoing

Green: complete

es-doc Earth System Documentation

is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

University of Reading

National Centre for Atmospheric Science NATURAL ENVIRONMENT RESEARCH COUNCIL

ES-DOC resources for CMIP6...

- Through the site and Further Info URL concept, ES-DOC supplies browser-based assets with defined routes to find standardised CMIP6 information systematically
  - project and site root page: es-doc.org

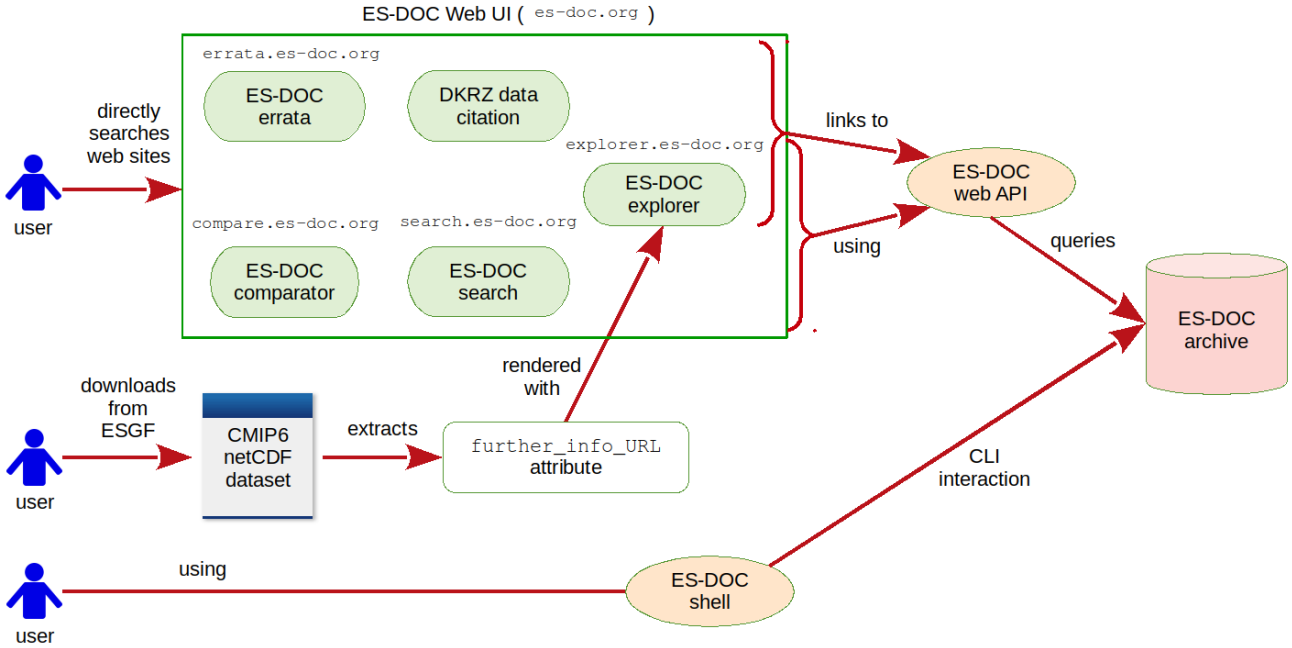
- Diversity of these entry points facilitates access:
  - website enables "top down" searching through, and exploration over, all available documentation...
  - ... whereas the Further Info URL binds it all together in a "bottom up" approach

… and how these may be useful

- A wide range of questions relating to CMIP6 can be probed using the ES-DOC resources. For example:
  - Which simulations were used in ensemble A?
  - Which models ran experiment B?
  - What details are given about component C of configured model D?
  - What are the relationships between model E and model F?
  - Which models produced the variable G?
  - Are there any known errors on dataset H?
  - What are the protocols on experiment I?
  - How can I formally cite dataset J?

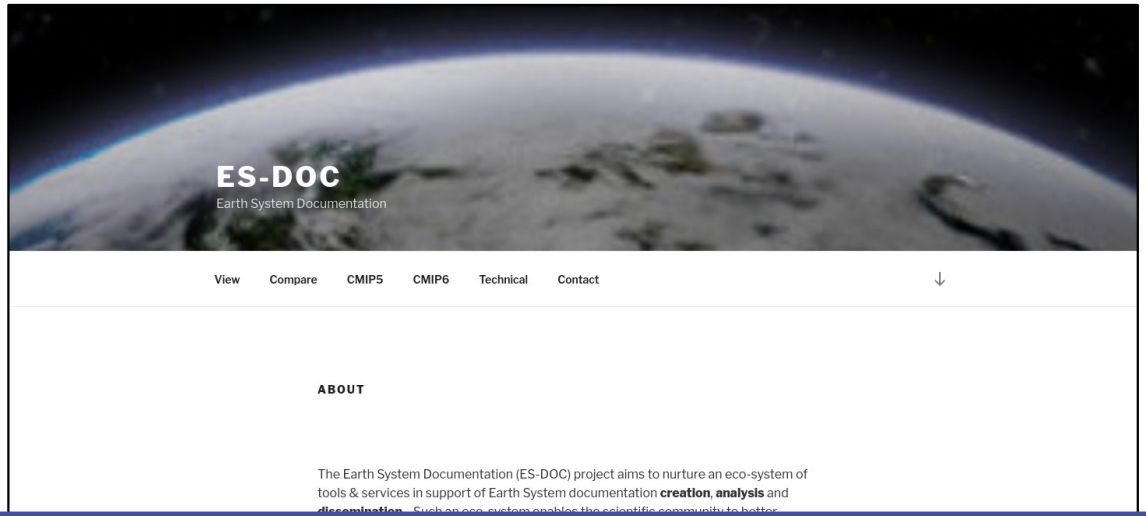
Overview of access routes

- Two online "portals" into ES-DOC for CMIP6 depicted
  - command-line route is mostly for developers and not discussed here

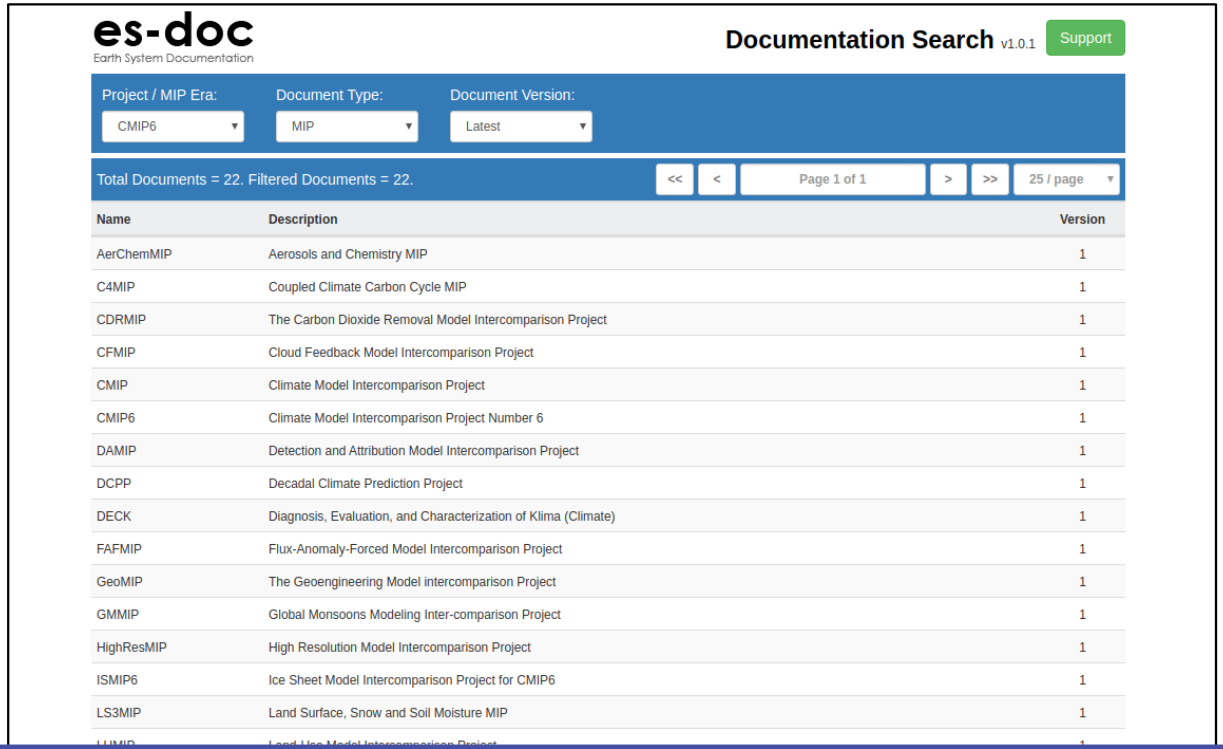


The ES-DOC website

- Via the site, can *search*, *view*, *explore* and *compare* CMIP6 documentation, plus *discover errata and citations*
  - project & site root page: es-doc.org
  - CMIP6 overview landing page: es-doc.org/cmip6



… search with search.es-doc.org



es-doc
Earth System Documentation

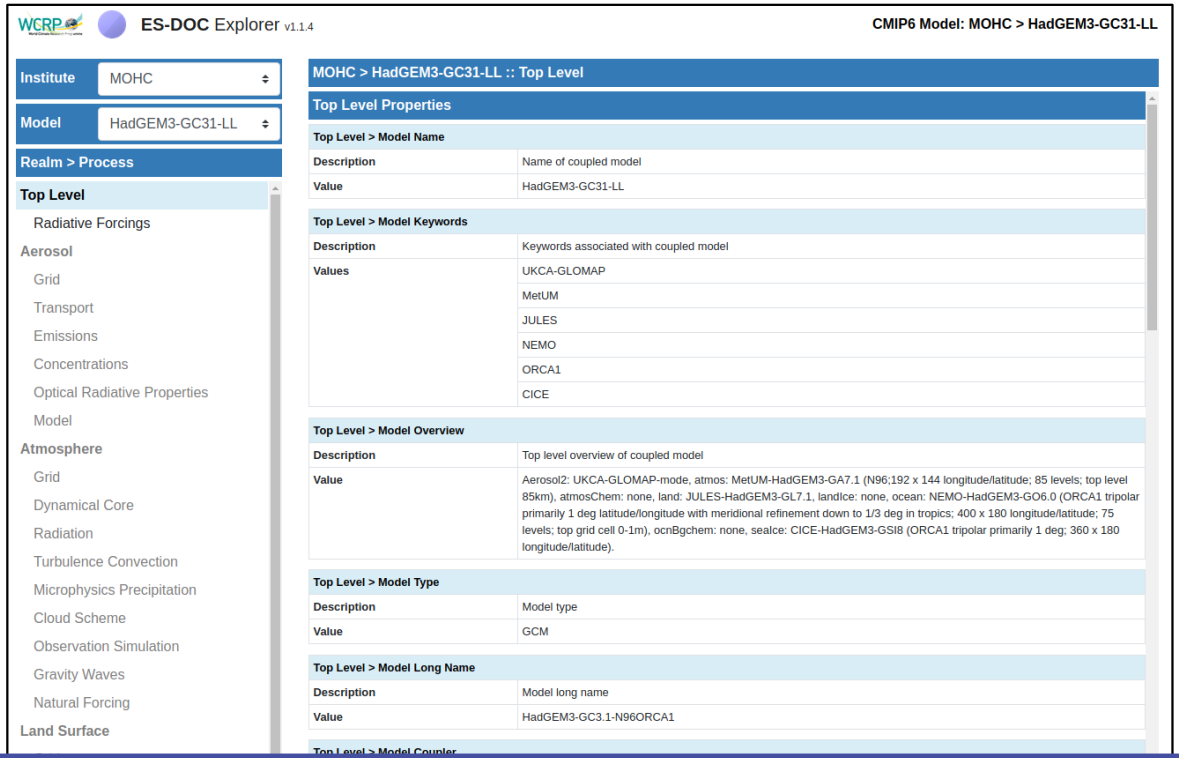
Documentation Search v1.0.1 Support

Project / MIP Era: CMIP6 | Document Type: MIP | Document Version: Latest

Total Documents = 22. Filtered Documents = 22. << < Page 1 of 1 > >> 25 / page

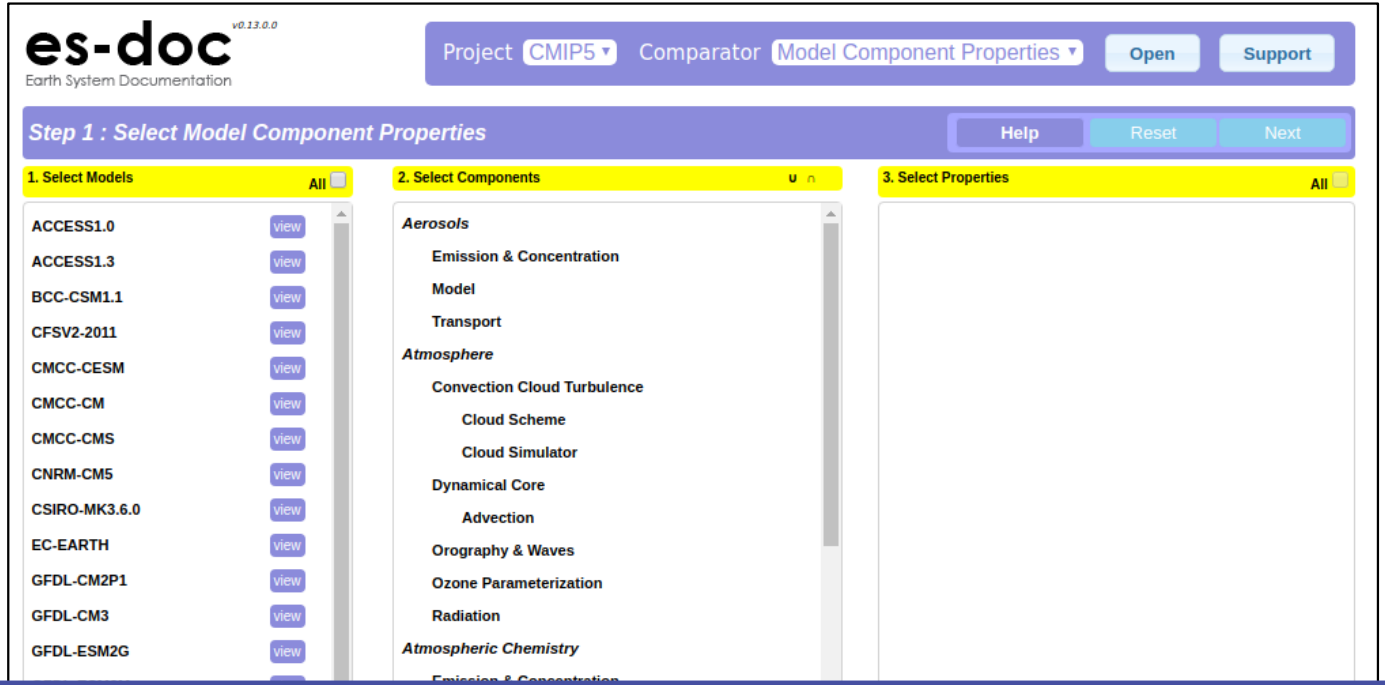
Name	Description	Version
AerChemMIP	Aerosols and Chemistry MIP	1
C4MIP	Coupled Climate Carbon Cycle MIP	1
CDRMIP	The Carbon Dioxide Removal Model Intercomparison Project	1
CFMIP	Cloud Feedback Model Intercomparison Project	1
CMIP	Climate Model Intercomparison Project	1
CMIP6	Climate Model Intercomparison Project Number 6	1
DAMIP	Detection and Attribution Model Intercomparison Project	1
DCPP	Decadal Climate Prediction Project	1
DECK	Diagnosis, Evaluation, and Characterization of Klima (Climate)	1
FAFMIP	Flux-Anomaly-Forced Model Intercomparison Project	1
GeoMIP	The Geoengineering Model intercomparison Project	1
GMMIP	Global Monsoons Modeling Inter-comparison Project	1
HighResMIP	High Resolution Model Intercomparison Project	1
ISMIP6	Ice Sheet Model Intercomparison Project for CMIP6	1
LS3MIP	Land Surface, Snow and Soil Moisture MIP	1
LUMIP	Land-Use Model Intercomparison Project	1

… explore with explore.es-doc.org



… compare at compare.es-doc.org

- Currently for CMIP5, coming later in 2021 for CMIP6



Web services

- Errata service*:
  - found at: errata.es-doc.org
  - discover/track known issues & amendments across published datasets
- Citation service†:
  - found (linked within but external to the ES-DOC site) at cmip6cite.wdc-climate.de (landing page) and bit.ly/CMIP6_Citation_Search (search application)
- Controlled vocabularies:
  - 'pyessv' utility library endpoint at: pyessv.es-doc.org
- Documentation creation and publication:
  - 'pyesdoc' utility library endpoint at: api.es-doc.org

* ESGF co-manage the errata service

† the citation service is hosted and managed outside of ES-DOC by DKRZ

More on errata

- An effective issue tracker platform for CMIP6
  - let's have a look around at errata.es-doc.org
  - described for example in this paper

- Aims to keep track of the issues affecting specific versions of datasets/files

- Designed and built on top of the **P**arent-**ID**entifiers handle service, ensuring maximum usability of the ESGF ecosystem and encapsulated in the ES-DOC structure

More on citations

- DKRZ Data Citation Service provides users of CMIP6 data with the *formal* citation that should accompany any use of the datasets.
- Addresses disconnect between data and citation
- **99%** of the registered CMIP6 data is citable, so cite the datasets you use!
- Let's explore the service from the landing page at cmip6cite.wdc-climate.de

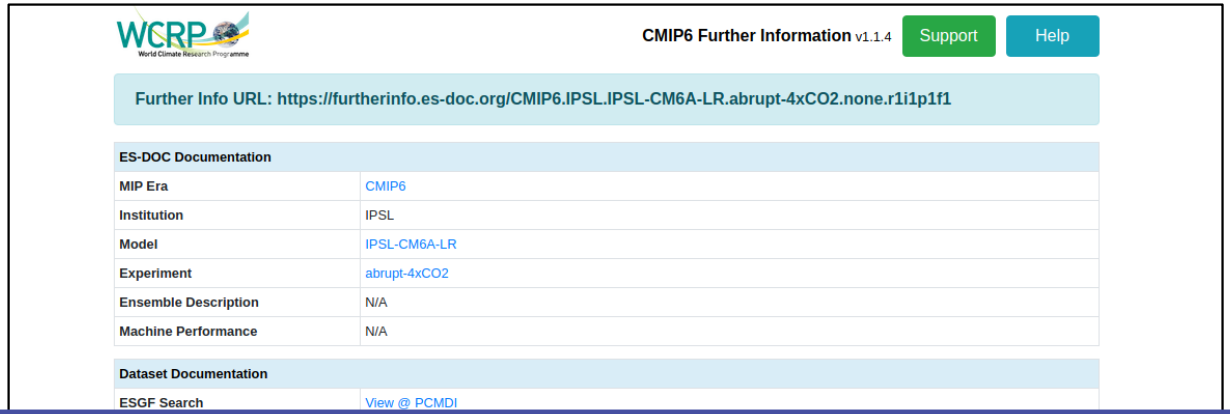
ES-DOC guidance pages

- ES-DOC site also has numerous pages dedicated to providing guidance and background information
  - root landing page for ES-DOC for CMIP6: es-doc.org/cmip6

- Notably, there are a set of screencasts (videos) explaining processes and the structure of CMIP6 and the role of ES-DOC for documentation and ESGF for data provision
  - the screencasts are embedded, with the corresponding dialogue also in text, at: es-doc.org/cmip6-screencasts

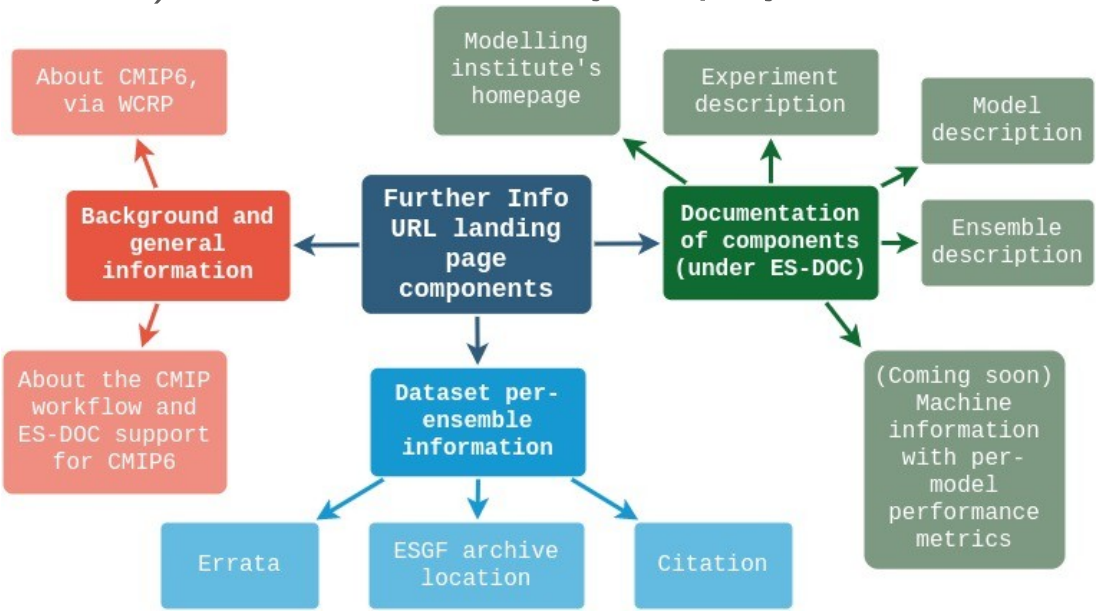
The Further Info URL

- *Further Info URL:* a landing page from which all CMIP6 documentation relevant to given data may be accessed
- The URL address format is systematically constructed
- More detail and a working example are outlined at: es-doc.org/cmip6-ensembles-further-info-url



Further Info URL as a page

- Each page is dynamic (auto-generated from available information) and a user-friendly display of numerous links:



Further Info URL on datasets

- A `further_info_url` property is defined in the header of *all* CMIP6 netCDF files! It is a netCDF global attribute.

- With your netCDF manipulation tool of choice you can access it and open it in a browser to see all metadata, e.g:
  - Command line (the `-h` "header-only" option and `grep` are optional):
    ```
    $ ncdump -h <filename>.nc | grep further_info_url
    ```
  - Python via cf-python (or equivalently with cfdm e.g. by substituting `cfdm` for `cf` in this snippet), minimal example with a field from a file:
    ```
    >>> import cf
    >>> f = cf.read('<filename>.nc')[0]
    >>> f.get_property('further_info_url')
    ```

ESGF dataset walk-through

- Let's run through an example of finding some datasets via the Earth System Grid Federation (ESGF) data nodes and grabbing and viewing the Further Info URL
  - For documentation and detail on ESGF data nodes, see for example: esgf.github.io/esgf-user-support/user_guide.html

- ESGF datasets in combination with ES-DOC-supplied documentation is a way to probe the lifecycle of CMIP6: data, metadata, processes, background etc.

To summarise...

- ES-DOC, going since ~2013 (IS-ENES funding) aims to enable the scientific community to better understand and utilise Earth system model data by nurturing tools and services in support of relevant documentation (workflow metadata) creation, analysis and dissemination
- ES-DOC operational for CMIP6 since December 2018
- There are two core browser-based entry points for accessing the published structured documentation:
  - a dedicated website with applications to search, view and explore
  - a Further Info URL concept attached to all CMIP6 netCDF files

Thanks, any questions welcome!

- ES-DOC website:
  - project & site root page: es-doc.org
  - CMIP6 overview landing page: es-doc.org/cmip6
  - search app (discover and view all documentation): search.es-doc.org
  - explorer (model formulations + further info URL): explore.es-doc.org
  - comparator: coming later in 2021 (for CMIP6) at compare.es-doc.org

- Further Info URL (get it from your CMIP6 netCDF files!):
  - about the concept, including a working example: es-doc.org/cmip6-ensembles-further-info-url