ES-DOC and CMIP source vocabularies

Initial Conversations around aligning ES-DOC and CMIP7 controlled vocabularies

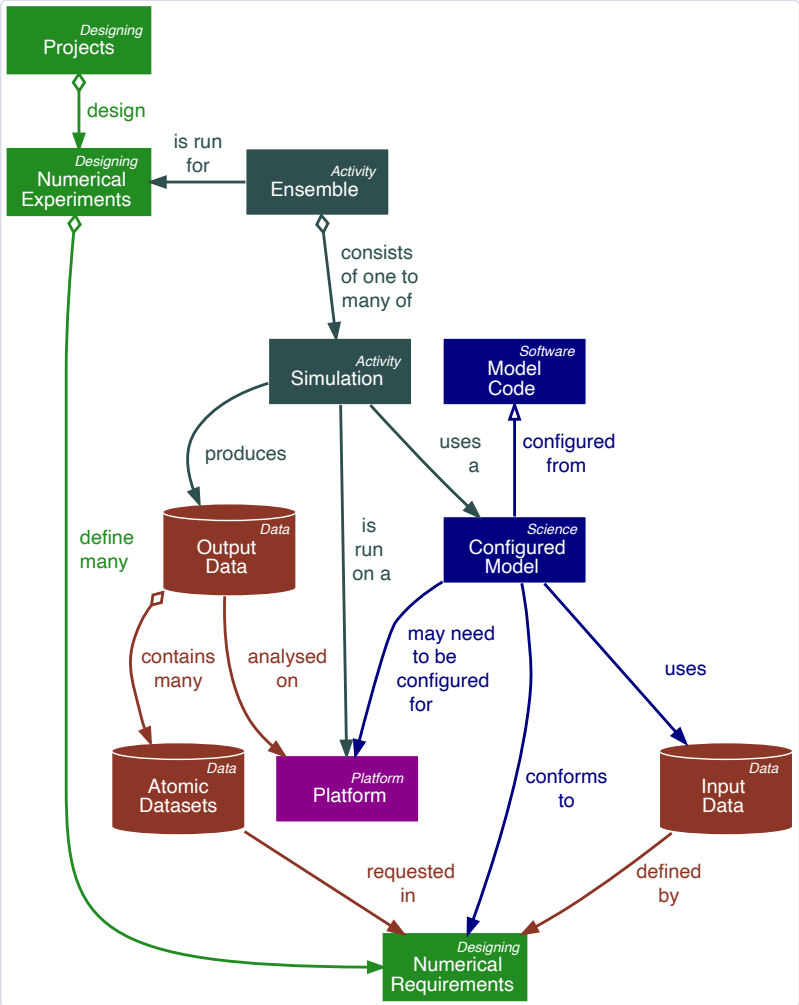
Bryan Lawrence, May 16, 2023
V1.1

Introduction

The context for this document is an attempt to align vocabularies so that

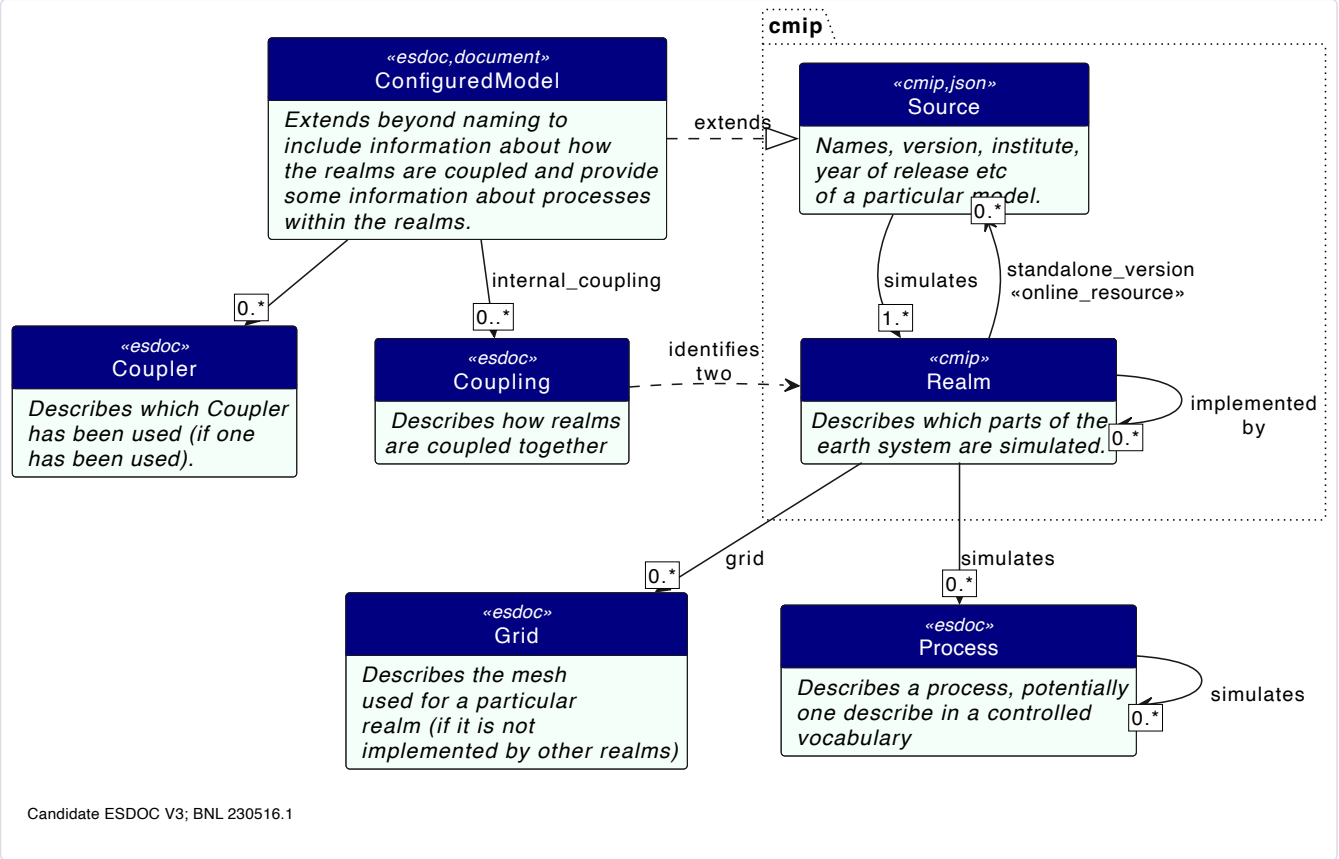
- the CMIP7 and ES-DOC7 information is consistent, and
- we minimise the number of times we ask for the same information.

We are here concentrating on the model documentation, i.e. the dark blue parts of the following depiction of the full workflow:



AltText

An an initial attempt to think about aligning the vocabularies appears in this diagram, more detailed attempts are described below.



The goal then, for today, is to consider what we might do so that the cmip7 source vocabulary is fit for being used in this way. In practice we will see that we need just a bit more controlling of what goes where in the json content held at PCMDI.

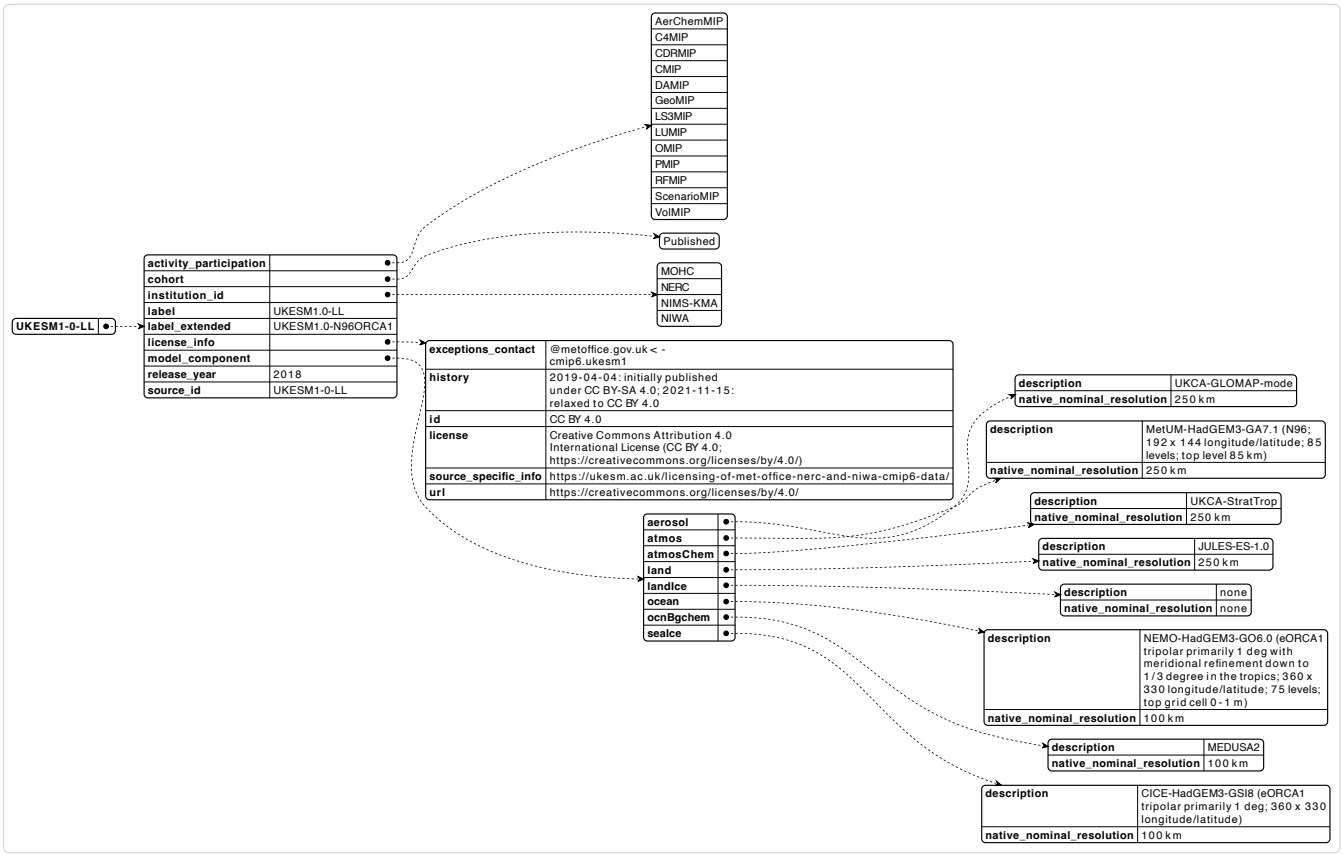
CMIP6 Source ID

The current CMIP6 source id controlled vocabulary consists of a set of model descriptions in json of the following form (note long lines are truncated here):

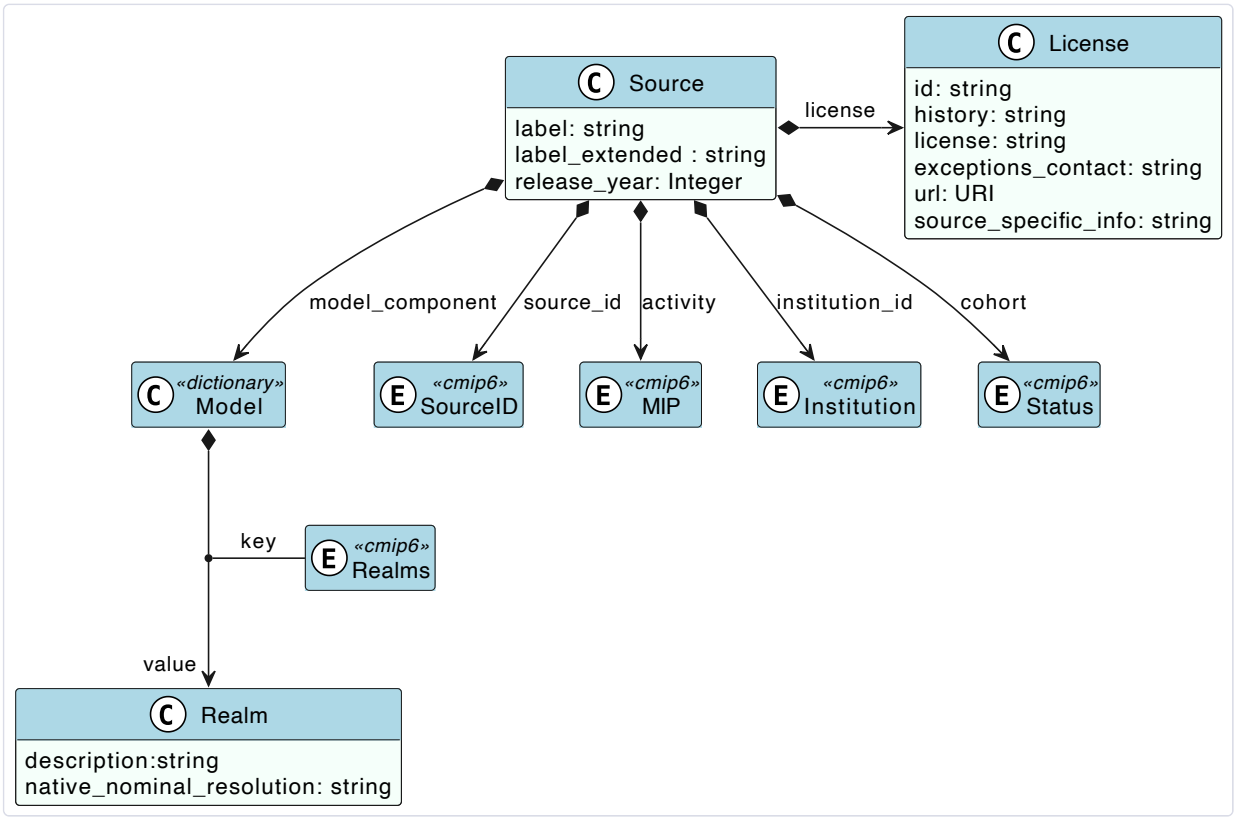
```language-json
"UKESM1-0-LL":{
    "activity_participation":[
        "AerChemMIP",
        "C4MIP",
        "CDRMIP",
        "CMIP",
        "DAMIP",
        "GeoMIP",
        "LS3MIP",
        "LUMIP",
        "OMIP",
        "PMIP",
        "RFMIP",
        "ScenarioMIP",
        "VolMIP"
    ],
    "cohort":[
        "Published"
    ],
    "institution_id":[
        "MOHC",
        "NERC",
        "NIMS-KMA",
        "NIWA"
```

```
    ],
    "label":"UKESM1.0-LL",
    "label_extended":"UKESM1.0-N96ORCA1",
    "license_info":{
        "exceptions_contact":"@metoffice.gov.uk <- cmip6.ukesm1",
        "history":"2019-04-04: initially published under CC BY-SA 4.0; 2021-11-15: rela
        "id":"CC BY 4.0",
        "license":"Creative Commons Attribution 4.0 International License (CC BY 4.0; h
        "source_specific_info":"https://ukesm.ac.uk/licensing-of-met-office-nerc-and-ni
        "url":"https://creativecommons.org/licenses/by/4.0/"
    },
    "model_component":{
        "aerosol":{
            "description":"UKCA-GLOMAP-mode",
            "native_nominal_resolution":"250 km"
        },
        "atmos":{
            "description":"MetUM-HadGEM3-GA7.1 (N96; 192 x 144 longitude/latitude; 85 l
            "native_nominal_resolution":"250 km"
        },
        "atmosChem":{
            "description":"UKCA-StratTrop",
            "native_nominal_resolution":"250 km"
        },
        "land":{
            "description":"JULES-ES-1.0",
            "native_nominal_resolution":"250 km"
        },
        "landIce":{
            "description":"none",
            "native_nominal_resolution":"none"
        },
        "ocean":{
            "description":"NEMO-HadGEM3-GO6.0 (eORCA1 tripolar primarily 1 deg with mer
            "native_nominal_resolution":"100 km"
        },
        "ocnBgchem":{
            "description":"MEDUSA2",
            "native_nominal_resolution":"100 km"
        },
        "seaIce":{
            "description":"CICE-HadGEM3-GSI8 (eORCA1 tripolar primarily 1 deg; 360 x 33
            "native_nominal_resolution":"100 km"
        }
    },
    "release_year":"2018",
    "source_id":"UKESM1-0-LL"
}
```


We can visualise that as



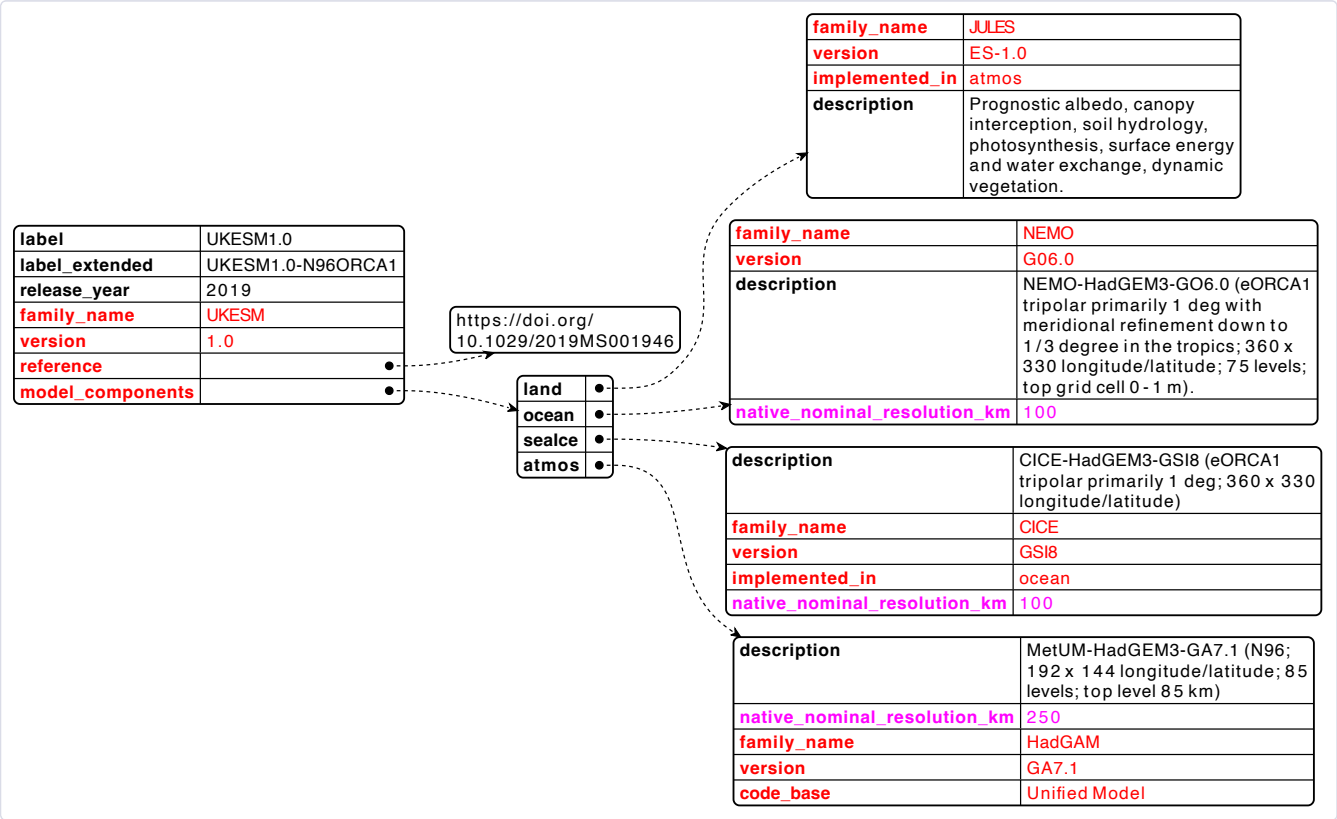
In "nearly UML" the structure looks like:



What changes would es-doc want? We'd want family names and version numbers in the `Source` and `Realm` and to modify Realm to know whether or not a Realm was implemented *as a* model component or *in_a* model component, and if it was *in_a*, which one? We would also like to control the resolution so the string number and units are forced to be an integer and 'km'. It would be *nice* to have a link to a source code repository.

(One might ask why there is a label attribute for a source, why is it different from the source_id, which itself is a label?)

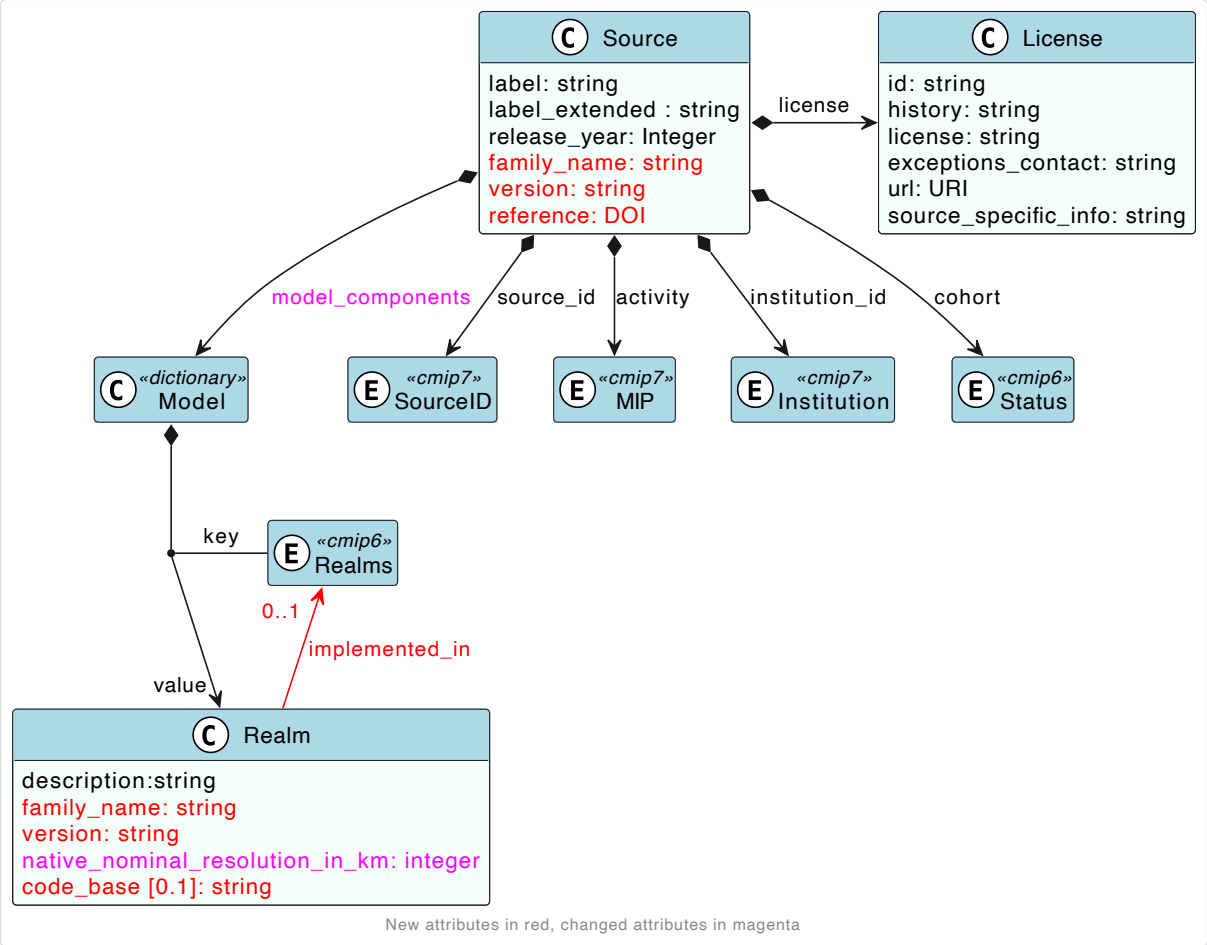
So, for an example familiar to some of us showing only some of entries and only a few realms (with truncated descriptions), UKESM1.0 would be:



Mind you, looking at what the MOHC chose as important description, it's mainly the mesh, so maybe we could pull the mesh out into it's own json element (i.e. pull the grid shown in the second diagram of this document into the cmip7 package). TO DISCUSS.

If the validation paper didn't exist at the time of data submission, we'd want to get it provided when available; es-doc could follow up on that.

The easiest way to achieve the core changes in source information collected at PCMDI would be to modify the the UML (and constraint the json) so as to get:



We could then pre-populate the es-doc description, and then request updates to get validation papers (which might follow publication of dat and data descriptions),

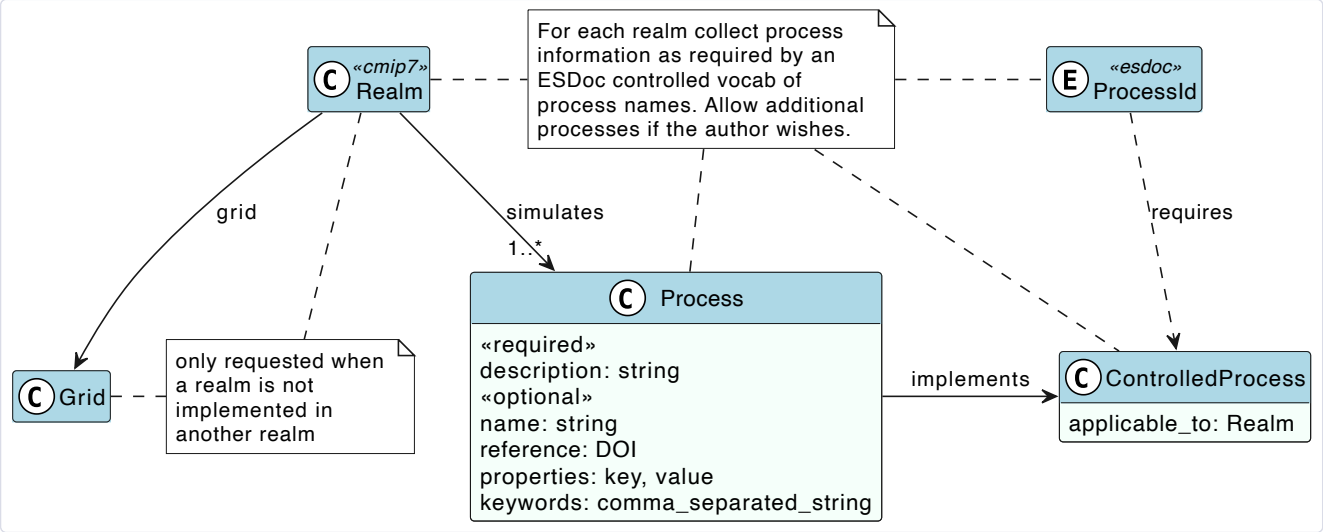
What then would ES-DOC do subsequently?

(This is not really in scope for the initial discussions today)

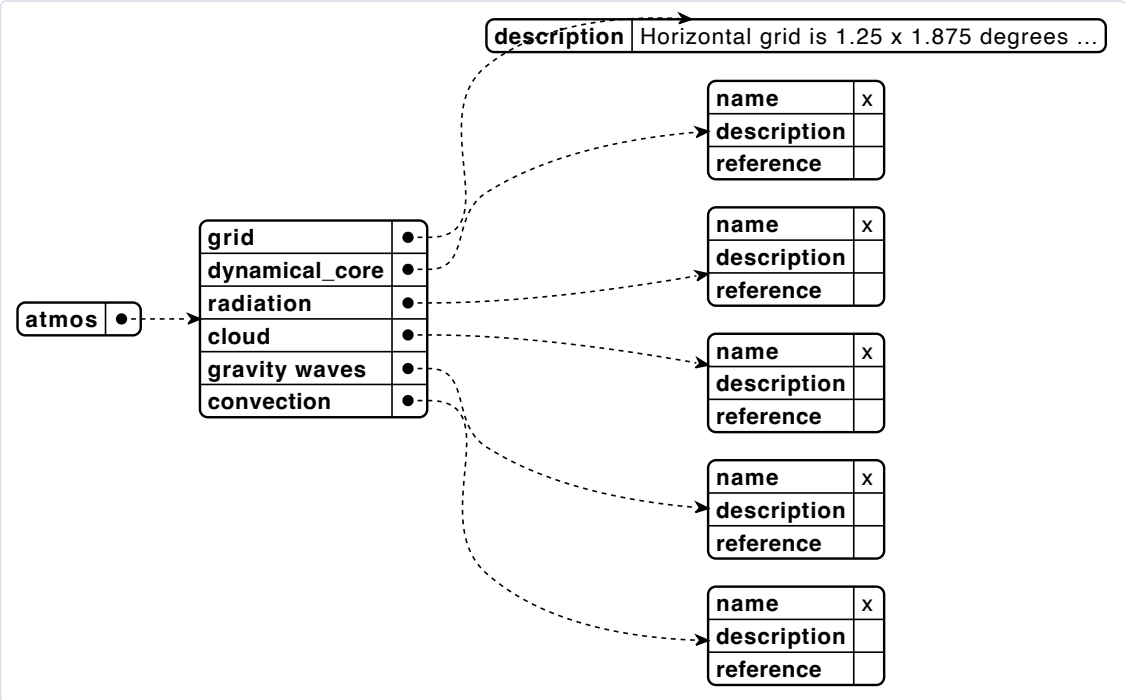
The only extra *Realm* information ES-DOC would then want to collect is some information about how the Realm was implemented in terms of key parameterisations. This sort of information has been collected since CMIP1 (see appendix), and it would appear that the most value comes with one layer down of information, exposing the major parameterisation groupings with a name, brief description, and a citation.

(Es-doc would also want to generate a coupling diagram so would need some coupling information.)

Then all the additional esdoc *realm_model information* would be described in the following UML:



We could prepulate a spreadsheet to collect this, with a sheet for each realm from the `<<cmip7>>` source_id vocabulary they would already have had to publish - or we could take json, e.g. for atmosphere the json might look something like this:



where the list of processes needs to be *controlled per realm* (i.e. belong to the ProcessID enumeration.)

We would allow all processes to include optional key, value properties and keywords, but not require them.

APPENDIX

A list of interesting historical model documentation datasets:

CMIP1:

- A list of models, with a paper for each of the coupled model, the atmosphere, and the ocean: cmip1 table1.
- How the models met experimental criteria (e.g. flux adjustment, initialisation): cmip1 table2.
- Ocean Model Documentation, covering dynamical scheme, vertical coordinate, eddy mixing , resolution: cmip1 table3

- Atmospheric Model Documentation, resolution, and a link to the documentation for the nearest AMIP model: cmip1 table4
- A table covering sea/ice land/ice properties (cmip1 table5) that links out to a block of text for each realm. Here, for example is a block of text for a particular sea ice realm (from the COLA coupled model elaborations):

> The sea ice parameterization (cf. the appendix to Schneider and Zhu 1998) is a simple prognostic single-layer thermodynamic model. Given surface fluxes provided by the atmospheric model, the scheme calculates changes in ice thickness and surface temperature, and the modified fluxes of heat and fresh water supplied to the underlying ocean. The time step is semi-analytic to maintain stability.
> Sea ice forms over the whole of an ocean grid box when the temperature of the topmost ocean layer is predicted to fall below the saltwater freezing temperature. Melting occurs at the top or bottom of the ice if the respective temperatures of these surfaces are predicted to be above this freezing point. Freshwater fluxes from the atmosphere are unmodified by the sea ice, but freezing or melting at the ice bottom causes appropriate changes in the freshwater flux to the ocean.
> The ice albedo is a constant, independent of surface temperature or ice melt. Solar radiation does not penetrate the ice and there are no other internal heat sources. Overlying snow cover also does not affect the ice thermodynamics. The temperature gradient within the ice is assumed to be linear, and the temperature at the bottom surface is prescribed to be the saltwater freezing point. The heat flux into the ocean is determined by conduction down the ice temperature gradient, which is controlled by the ice thickness and the top surface temperature. Heat is exchanged between the ice and the topmost ocean layer, consistent with freezing/melting, to keep the ocean surface temperature at the saltwater freezing point.
> Sea ice dynamics and rheology are not represented.