es-doc for cmip6

(not a summary of es-doc per se)

Bryan Lawrence, Sebastien Denvil, Mark Greenslade, Eric Guilyardi, David Hassell, Sylvia Murphy, Charlotte Pascoe, Allyn Treshansky

The es-doc consortium; IS-ENES2 and NOAA partners

17 February 2016

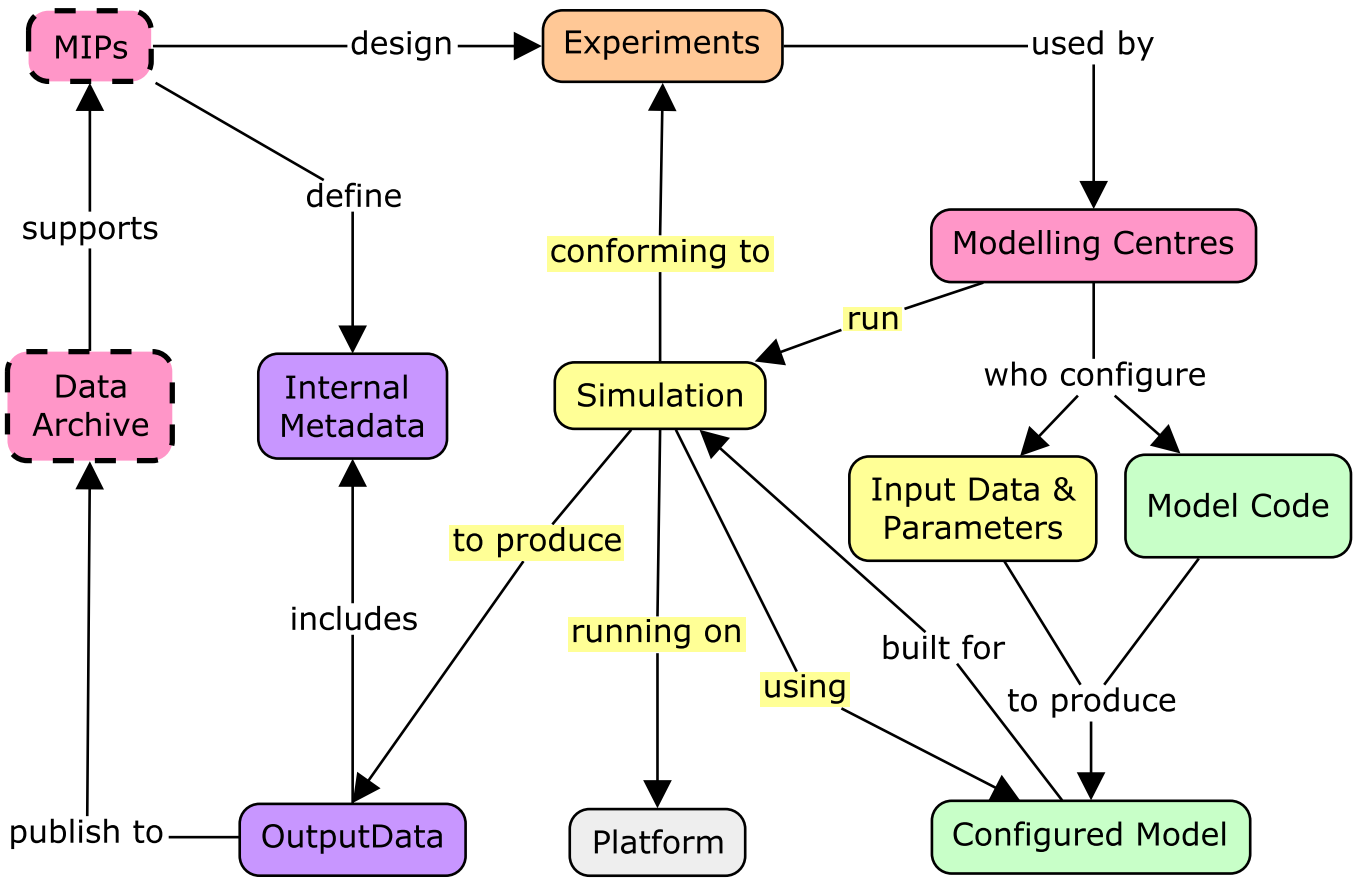
Definitions

- An experiment is an activity aimed at addressing a specific scientific problem.
- We formally describe such an experiment by means of the **NumericalExperiment** which describes the experimental aim, and is composed of a set of **NumericalRequirement**s which need to be met to address the experimental aim, these include any spatio-temporal constraints (what domain is simulated, for how long), forcing constraints (e.g. whether a historical or future scenario is used for anthropogenic emissions of radiatively important gases) etc.

Definitions

- An experiment is an activity aimed at addressing a specific scientific problem.
- We formally describe such an experiment by means of the **NumericalExperiment** which describes the experimental aim, and is composed of a set of **NumericalRequirement**s which need to be met to address the experimental aim, these include any spatio-temporal constraints (what domain is simulated, for how long), forcing constraints (e.g. whether a historical or future scenario is used for anthropogenic emissions of radiatively important gases) etc.
- A **Simulation** is a run of a configured **Model** which conforms to the **NumericalRequirement**s, runs on a **Platform** and produces output **Dataset**s.

Big Picture Workflow



Evolution

- CIM 1.x used for CMIP5; based on XML schema and XML documents produced by the "metafor questionnaire".
  - Limited set of "documents", with much manual work to create model and simulation descriptions)
  - (A document is intended to be a thing created and owned by one entity in the workflow.)

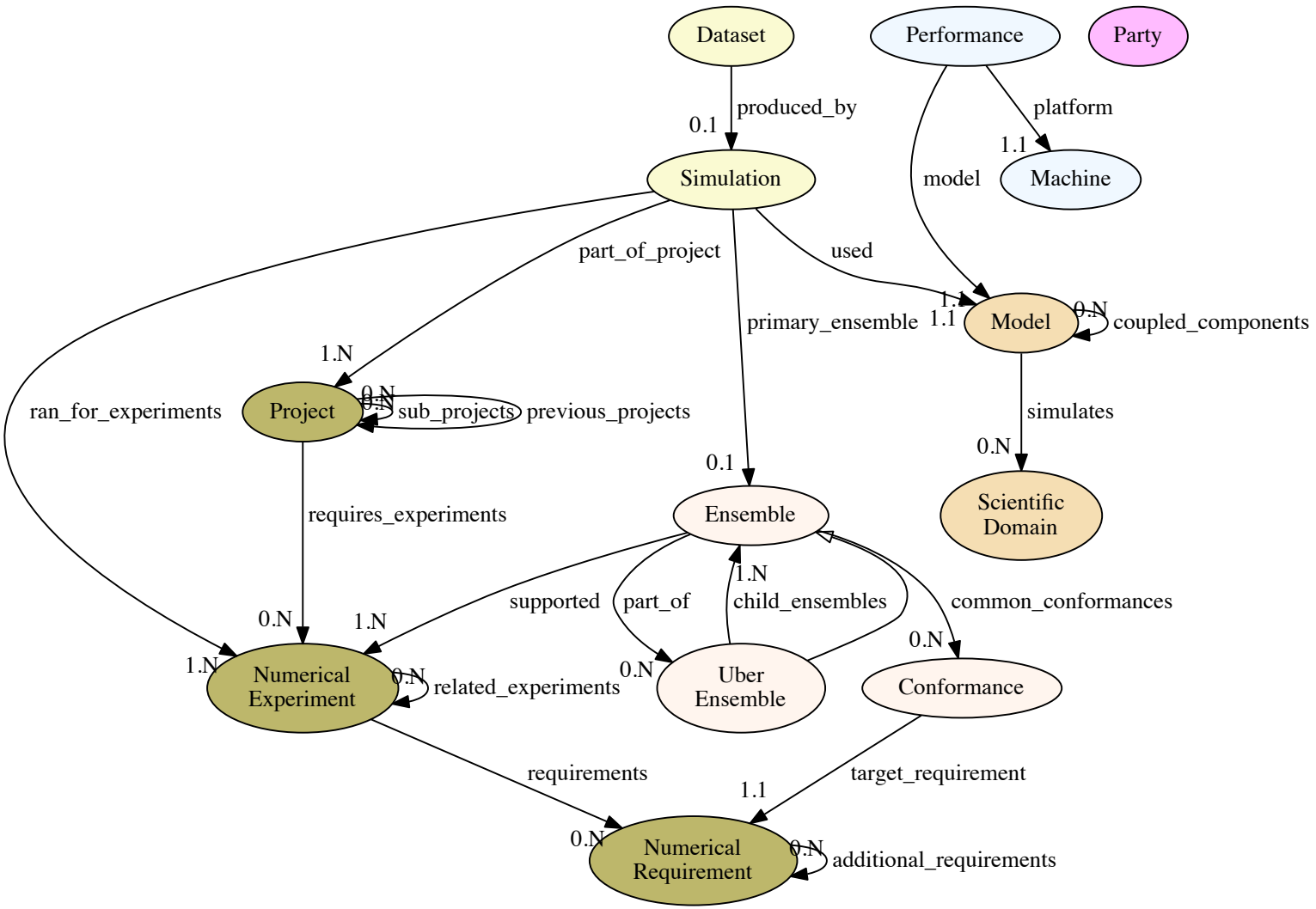
Evolution

- CIM 1.x used for CMIP5; based on XML schema and XML documents produced by the "metafor questionnaire".
  - Limited set of "documents", with much manual work to create model and simulation descriptions)
  - (A document is intended to be a thing created and owned by one entity in the workflow.)
- Many issues raised with CIM1.x and what we did for CMIP5 ...
- CIM 2.x to be used for CMIP6, based on a completely new method of defining the CIM
  - ... designed by **Mark Greenslade** in the toolchain for CMIP5.

Evolution

- CIM 1.x used for CMIP5; based on XML schema and XML documents produced by the "metafor questionnaire".
  - Limited set of "documents", with much manual work to create model and simulation descriptions)
  - (A document is intended to be a thing created and owned by one entity in the workflow.)
- Many issues raised with CIM1.x and what we did for CMIP5 ...
- CIM 2.x to be used for CMIP6, based on a completely new method of defining the CIM
  - ... designed by **Mark Greenslade** in the toolchain for CMIP5.
- Broad concepts are the same for CMIP6 as CMIP5, but there is more detail.
  - Considerably more "documents", and ways of creating documents.
  - Fixes known problems identified in CIM1.x and addresses community requirements.

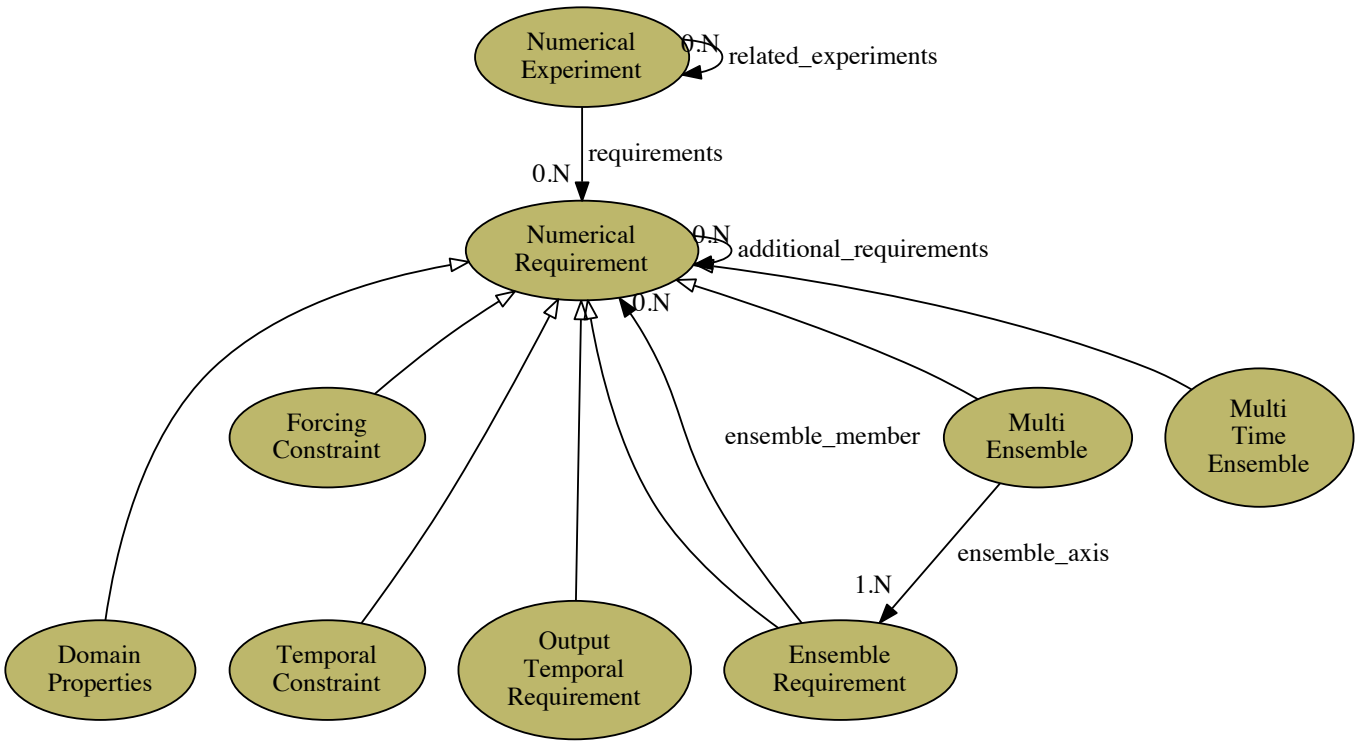
In more detail: CIM2 classes for CMIP6



CMIP6 Phases and Documents

1. Describing the experiments and their requirements so that simulations (and their models) are correctly configured.
2. Documenting the actual model configurations used and what scientific processes those models encapsulated.
3. Acquiring information about which simulations were run and what data is available.
4. Linking simulation information to the various experimental requirements including information about how the simulations might have varied along ensemble axes (conformance).
5. Recording the amount of computing used to deliver the experiments (to support assessments of future computing requirements), and
6. Documenting the data products themselves.

Experiment classes



End-to-End System

- A complete toolchain for producing experiment descriptions via spreadsheets through to display at es-doc.org has been built and exercised (Charlotte and Mark).
- Charlotte is currently fully documenting CMIP6 experiments with the aim to get these included in CMIP6 documentation.
- This all uses an early branch of CIM2, and will be upgraded to the latest version **only when it has been completely ratified via the modelling groups** - that is, content is dominating over tools . . . for experiments, and for now.

es-doc Earth System Documentation

Documentation Viewer v0.9.3.1 Support

CMIP6-DRAFT Experiment : SSP1-26

Overview

Project	CMIP6-DRAFT
Name	ssp1-26
Long Name	SSP1-2.6
Description	What: SSP-based RCP scenario with low radiative forcing by the end of the century. Following approximately RCP2.6 global forcing pathway with SSP1 socioeconomic conditions. Radiative forcing reaches a level of 2.6 W/m2 in 2100. Concentration-driven. Why: The scenario represents the low end of the range of plausible future pathways. The scenario depicts the "best case" future from the sustainability perspective.
Keywords	Scenario | SSP | RCP | SSP1 | RCP2.6 | future | climate change | IPCC | ScenarioMIP | Low | SSP-based RCP | Tier 1
Related Experiments	historical

Temporal Constraints

2014/01/01- 2100/01/01

Start Date	2014-01-01
Required Duration	86 years
Description	Scenario, from 2014 to the end of the 21st century.
Conformance Requested ?	False
Keywords	future | scenario | 2014 | 2100

Ensemble Requirements

Historical Initialisation

Name	HistoricalInitialisation
Description	What: Initialisation is from the end of the Historical experiment. Why: to provide continuity between simulations of the recent past and future scenario simulations.
Ensemble Type	Initialisation Method
Minimum Size	1
Conformance Requested ?	True
Keywords	initial conditions | initialisation | historical | scenario

Single Member Ensemble

Name	SingleMemberEnsemble
Description	One ensemble member
Ensemble Type	Initialisation
Minimum Size	1
Conformance Requested ?	False
Keywords	Single | simulation | run | ensemble

Forcing Constraints

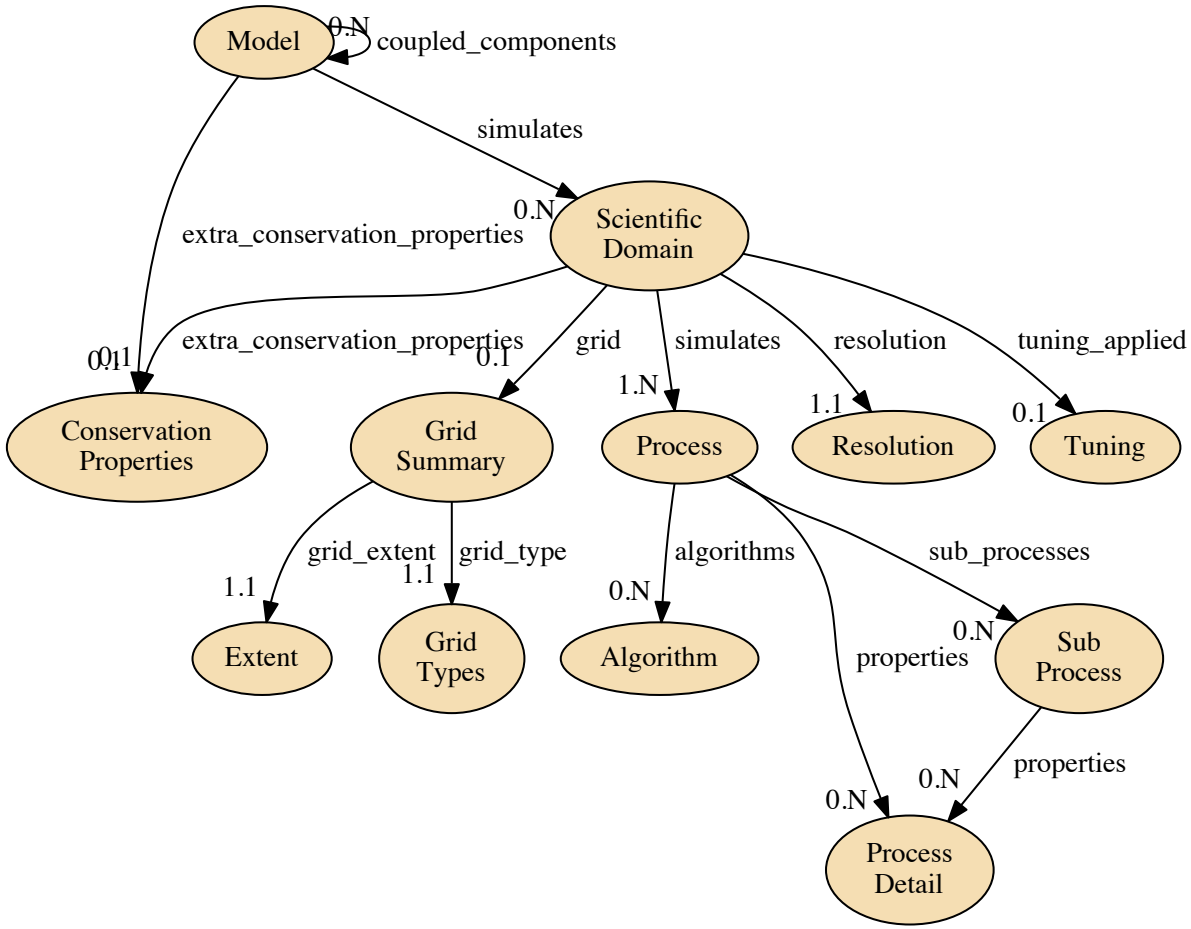
Representative Concentration Pathway 2.6 Aerosol Precursors

Name	RCP26aerpre
Description	What: Impose changing concentrations of RCP2.6 aerosol precursors. Why: Represents the low end of the range of plausible future pathways.

Key changes for describing Configured Models

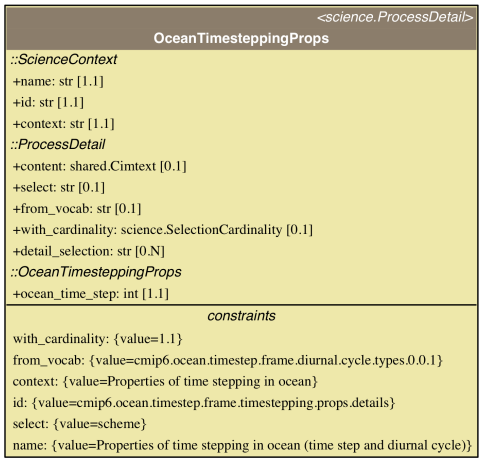
- CIM2 no longer confuses software components with scientific content (aka components)
- CIM2 is much simpler for describing scientific content, with no capability for infinite nesting.
- Design now aimed at simpler entry via spreadsheets (if desired) and easier configuration of comparison tables (conceptually, the software was already able to cope).
- The old complex vocabularies (not to be confused with lists of terms) are now defined using the same formalism as CIM2 itself (albeit with a shorthand notation).
- Everything can now be versioned and managed in Git

Scientific Descriptions



(some minor changes are still underway)

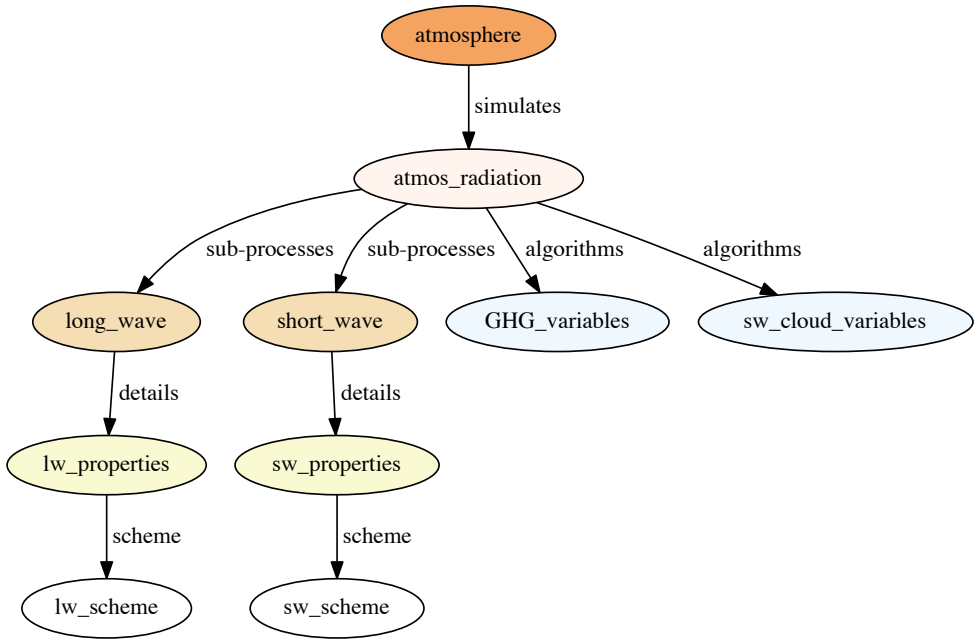
Some Vocabulary Examples



(cmip6 constrained version of science.ProcessDetail)

content	Free text description of process detail (if required).
select	Name of property to be selected from vocab
from_vocab	Name of an enumeration vocabulary of possible detail options.
with_cardinality	Required cardinality of selection from vocabulary
detail_selection	List of choices from the vocabulary of possible detailed options.
name	The name of this process/algorithm/sub-process/detail
id	Identifier for this collection of properties
context	Scientific context for which this description is provided
ocean_time_step	Ocean time step in seconds

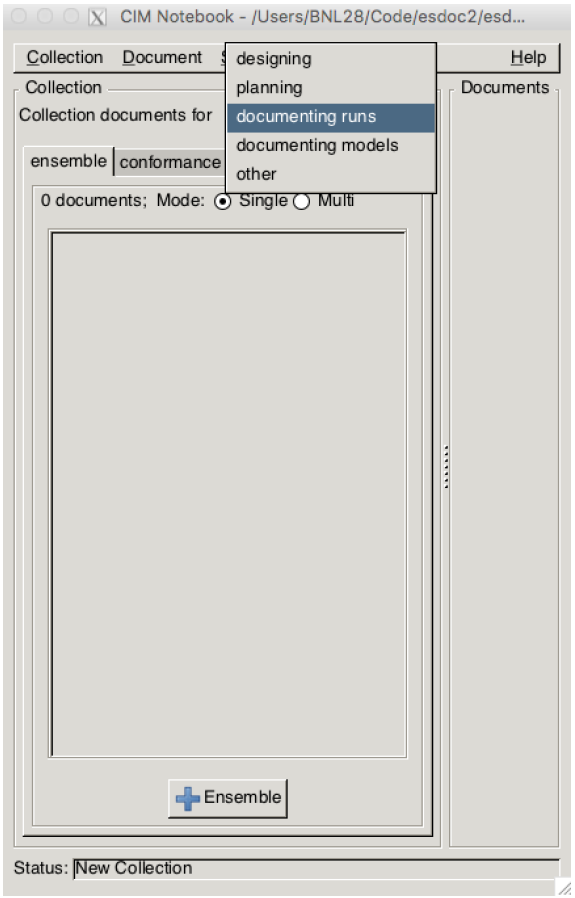
OceanTimesteppingProps



cmip6 (from atmosphere)

(All these figures autogenerated from the definitions.)

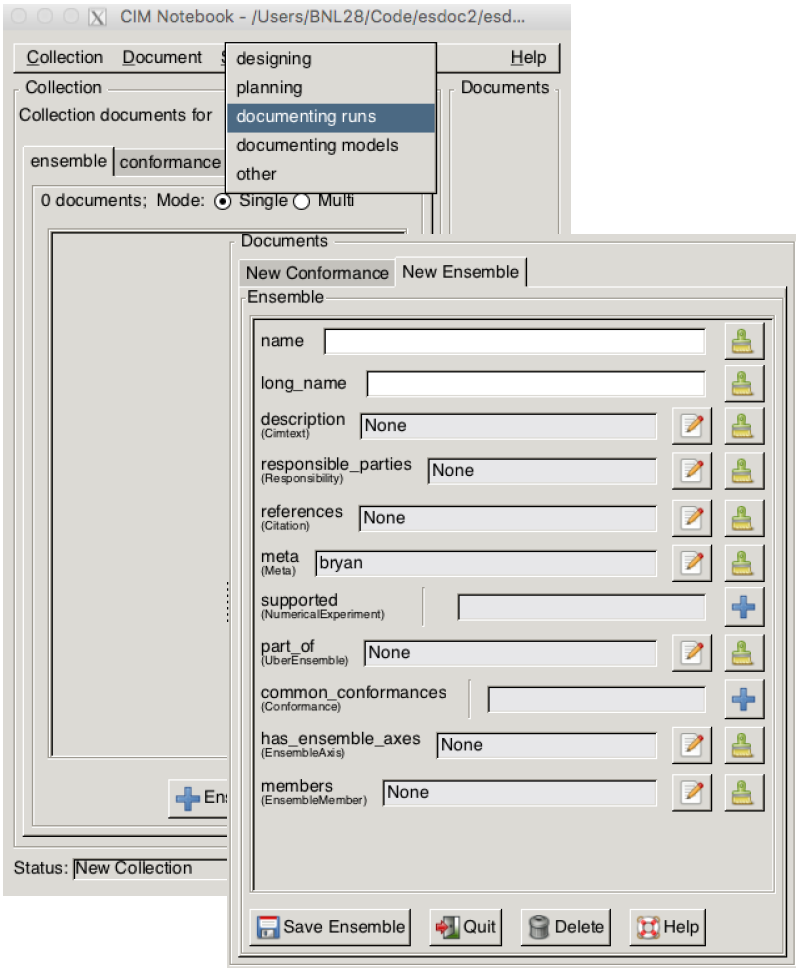
Most tooling now autogenerated



Developed a GUI desktop tool as part of the design of CIM2, which provides a complete data entry functionality, and can be used for the Milestone 113.

This tool is not intended for primary data entry by modelling groups, who we expect to use other routes, of which more later

Most tooling now autogenerated



Developed a GUI desktop tool as part of the design of CIM2, which provides a complete data entry functionality, and can be used for the Milestone 113.

This tool is not intended for primary data entry by modelling groups, who we expect to use other routes, of which more later

Managing notebook and CIM2

- As of **today**, the authoritative CIM2 is `https://bitbucket.org/bnlawrence/esdoc-nb/`.
- It will move to `https://github.com/ES-DOC/esdoc-mp/` as soon as is practical (hopefully by, or during the March es-doc sprint.
- At which point I have to stop being in the critical path ...

Everything else (which is much more)

Sustained effort by Mark Greenslade to ensure that the CIM2 developments will be supportable within the es-doc website and toolchain. Key components will include (but not be limited to):

1. esdoc-py-client: python tools for creating and manipulating documents (and other things)
2. esdoc-shell: command line shell tools for es-doc
3. esdoc-web: software for the esdoc website.
4. esdoc-mp: the "canonical" meta-programming framework
5. esdoc-api: web service API in support of ES-DOC eco-system
6. esdoc-js-client: tool for calling esdoc from javascript

Also major effort by Allyn Treshansky:

1. esdoc-questionnaire: tooling for creating documents using a traditional questionnaire technique.

Documents and Status

Document Family	Creation Tools	Point of Contact	Status
Experiments	Spreadsheet, notebook	Charlotte Pascoe	Complete pipeline exists
Science	pyesdoc, NOAA questionnaire, (notebook)	Bryan Lawrence (David Hassell)	In development
Dataset (inc Simulation)	python parser (in esgf publication workflow)	David Hassell	tested for cmip5 awaiting cmip6 file finalisation
Ensemble (further info)	NOAA questionnaire, Spreadsheet	tbd	yet to start tools
Party, References	NOAA questionnaire, uploader, (zotero connector?)	tbd	yet to start tools
Performance	NOAA questionnaire, spreadsheet, notebook	tbd	yet to start tools

All documents will be viewable in es-doc.org, although there is a significant amount of work still needed on templates. It will be possible to download tables and documents for use in papers and presentations.

Timeline for Scientific Vocabularies

Time line 2016	Activity or milestone	Owner
By now	Identify experts to produce first version of realm + WGCM tables	Eric and David
February	Finalise first version of ocean/seaice/atmosphere + WGCM tables	Charlotte, Bryan, Eric
F2F (March)	Finalise CIM2 and CMIP6 documentation workflow	ES-DOC
End March	First version of (all) realms and WGCM tables ready (taking into account inputs from core experts)	Charlotte, Bryan, Eric + experts for other realms
April-June	Review of CV.py by modelling groups and users	David and Eric
End June	CV.py files + es-doc tools (creation and view) ready for wider community testing	ES-DOC/WIP
July	Release WGCM model tables	ES-DOC/WIP
Dec	Release CMIP6 documentation complete procedure for community	ES-DOC/WIP

Putting it all together

There is a draft WIP paper on this which needs to go onto a public git repository soon.

Issues remain with relationship with internal file metadata in CMIP6 and with the links between content in file and es-doc ....

We have a major sprint/workshop planned for March 8/9/10 in Abingdon, UK, at which point most of the major decisions will be finalised, and we will finalise our plans for IS-ENES2 deliverables and milestones.

The Future

- **es-doc** relies on key competencies:
  - Scientific direction, maintenance, management
  - Software leadership, project management, support for developers.
  - Software engineering
  - Systems
  - Support for producers and consumers
- . . . but it will have no core funding anywhere after the end of IS-ENES2.

The Future

- **es-doc** relies on key competencies:
  - Scientific direction, maintenance, management
  - Software leadership, project management, support for developers.
  - Software engineering
  - Systems
  - Support for producers and consumers
- . . . but it will have no core funding anywhere after the end of IS-ENES2.

Currently implemented in:

- Too few people, who are part-time, and not always able to work on the same things at the same time.
- Too dependent on Bryan and Mark (in different ways).

This is only going to get worse if we can't get

- **New funding**, that is
- Capable of supporting at least the Bryan and Mark **functions** full time with some backup elsewhere.