

Climate Model Simulation Metadata at the MetOffice

IS-ENES-2 Workshop: Meta-data Generation During Experiments

Mark Elkington, 21 January 2014



Topics

- MOHC Metadata Repositories and Workflow
- CMIP5 metadata production
- Planned CMIP6 era metadata production
- Required MIP Infrastructure Changes



MOHC Metadata Repositories

Science Wiki pages

CREM

Climate Research Experiment Management

SQL database system used to manage "science" metadata from MOHC models and major experiments

MOOSE

Met. Office Operational Storage Environment

SQL database system used to manage "data" metadata for climate model files held in the Met Office mass storage system

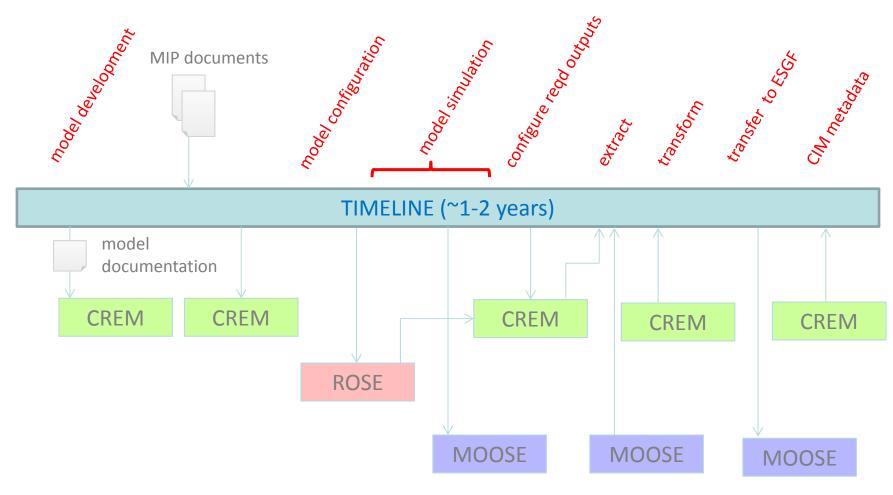
ROSE

Model Configuration Environment

SQL database system for managing model configurations and simulation configurations

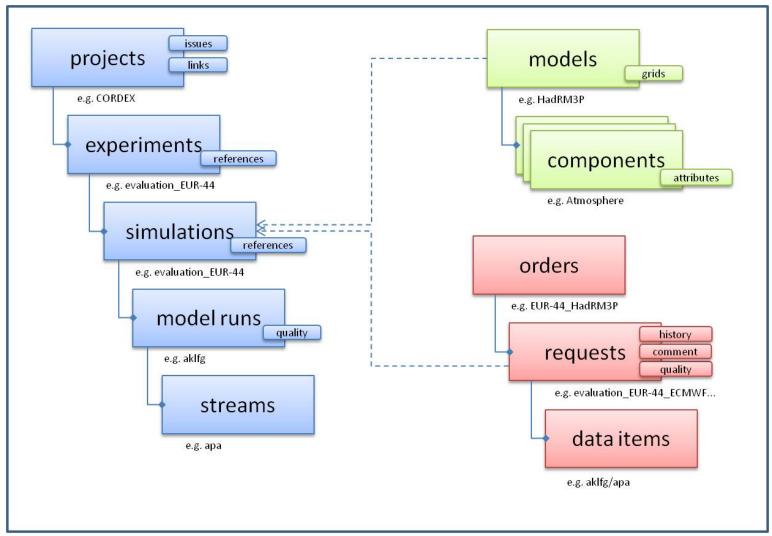


MOHC Metadata Work Flow





CREM schema overview

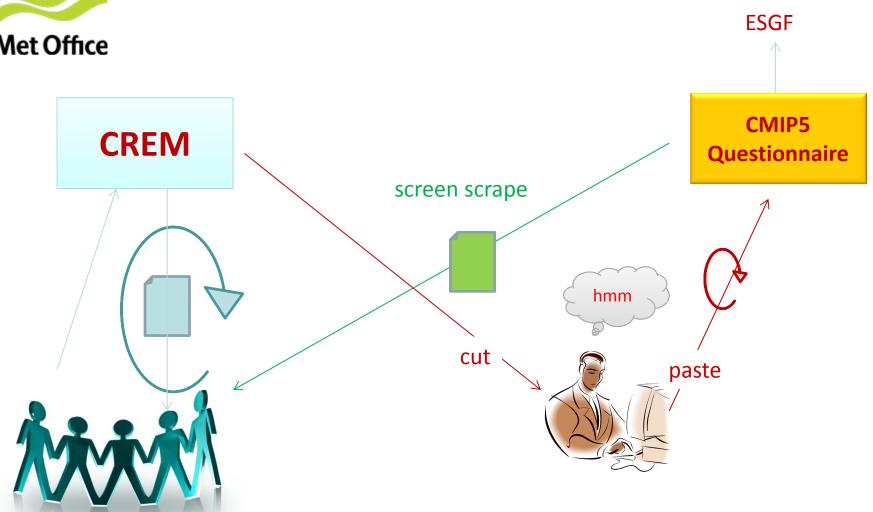




CMIP5 Metadata Experience



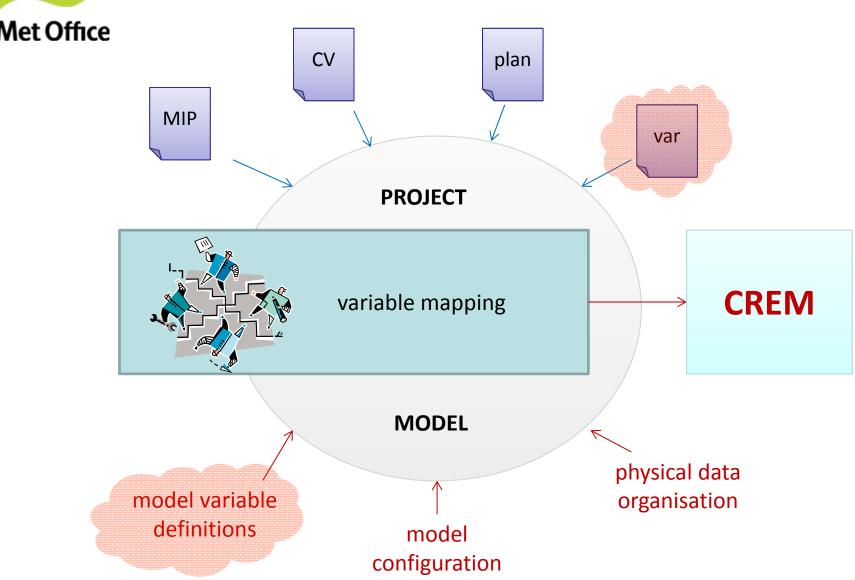
CMIP5 Metadata Creation - CIM



Science Teams across UK



project supplied metadata





CIM Metadata

Good

- Good attempt at adding structure to metadata
- Publishing model
- Recent ES-DOC tools show promise
- Resource for the future
- Lots of lessons learned

Not So Good

- Information of little value to our scientists – esp. experiment info.
- Questionnaire not a good model to enter complex metadata
- Analysis/presentation tools were too late
- Opaque versioning



Experiment Metadata

Good

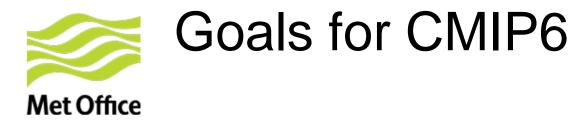
- MIP data good
- CV data OK but may not work well for multi MIP environment

Bad

- "Required output" data hard to use without lots of manual interpretation
- Experiment plan format not suitable to establish explicit conformance 'requirements'



CMIP6 Metadata Experience?



- Data production configuration driven more directly by project supplied metadata – more structured outputs required from experiment planning process.
- CIM outputs created directly from CREM metadata is created as direct output of scientist model configuration.
- CIM outputs published directly to ESGF will require effective verification tool, allowing flexibility where possible.



Required Output Scheme

Project (e.g. cordex, cmip6, geomip)					
	experiment group (e.g. rcp, historical, decadal, evaluation etc.)				
		Fre	Frequency (e.g 3h, 6h, daily, monthly etc.)		
			Variable (includes attributes such as start, end, vertical grid, ensemble (include/exclude), experiment (include/exclude)		
			variable		
			variable		
	frequency				
experiment group					



Infrastructure Enhancements Wishlist (metadata related)

- Required Outputs in structured format (XML/JSON) and held under configuration control with MIP tables.
- Direct publishing of CIM documents use of questionnaire interface not a requirement
- Version Management of CIM documents
- CIM usage and tools review to improve relevance to MIP science community
- Mandatory Quality Control checks to be specified explicitly at the same time as the data format specification is published.
 Minimise checks required.

