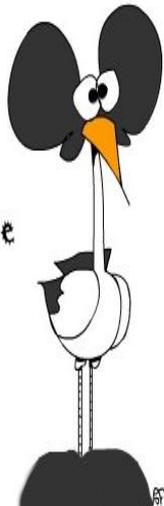


METAFOR

et les métadonnées CMIP5

MéTafOr



Orthophonie

Strasbourg



**Le projet METAFOR
et CMIP5**

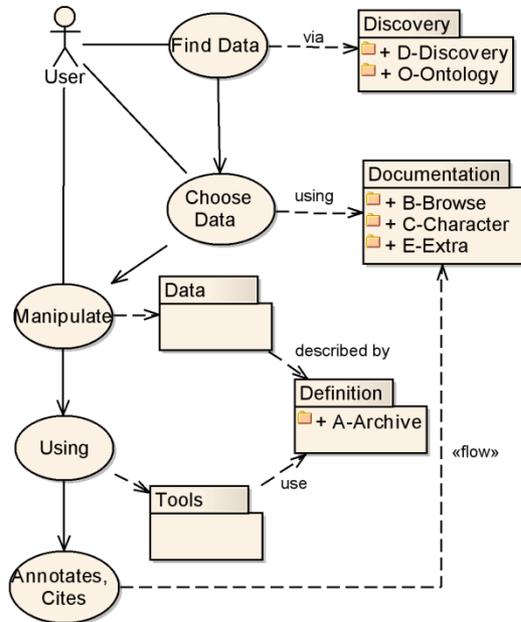
Common Metadata for Climate Modelling Digital Repositories

- Projet européen (EU-FP7)
- Initiative du European Network for Earth System Modelling (ENES)
- mars 2008- février 2011
- 11 puis 12 partenaires



- NCAS, University of Reading, UK
- BADC, STFC, UK
- CERFACS, France
- Models and Data, MPI, Germany
- IPSL, CNRS, France
- University of Manchester, UK
- Met Office, UK
- ANM, Romania
- Météo France, CNRM, France
- CLIMPACT, France
- CICS, Princeton University, USA
- Universidad de Cantabria, Spain

A quoi s'intéresse METAFOR ?



Accès, manipulation et utilisation des données issues de la modélisation climatique

- **préservation/longévité** (archives numériques)
- **interopérabilité** des système de distribution/exploitation
- **comparabilité** (exercices multi-modèles)
- **qualité/conformité** aux spécifications d'un exercice
- **ré-employabilité** par d'autres secteurs d'activité



Il faut **documenter** ces données Climat de façon **précise et univoque**

"Tous" de la même façon

Pas seulement ce qu'elle sont, mais aussi **comment** elles ont été produites

**Standard de
Métadonnées***

- Identification
- Accès
- Compréhension

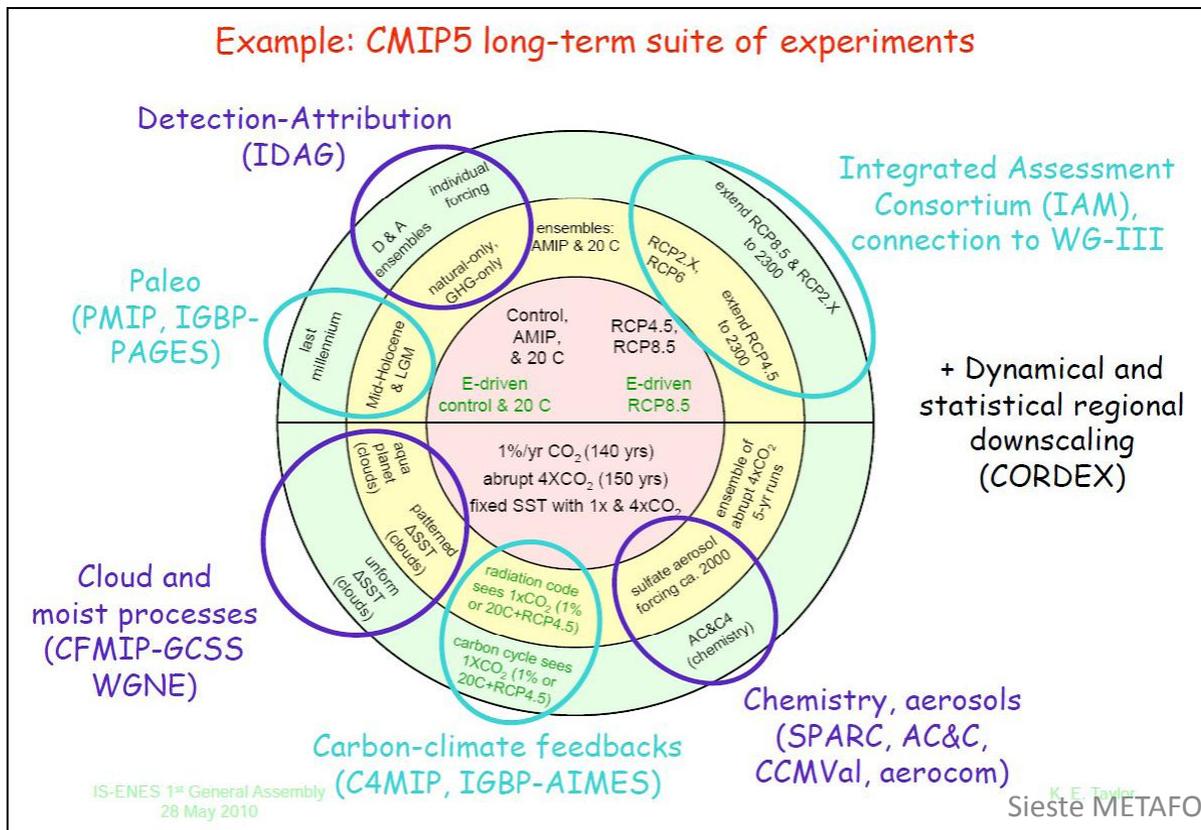
*Une métadonnée est une information (une donnée) caractérisant les données.

Au service de l'exercice CMIP5

CMIP5 (IPCC-AR5, 2013): Coupled Model Intercomparison Project Phase 5

- ~20 groupes de modélisation du climat
- ~30 configurations de modèles
- ~60 expériences numériques
- ~90 000 ans de simulation
- Large spectre de questions scientifiques adressées...

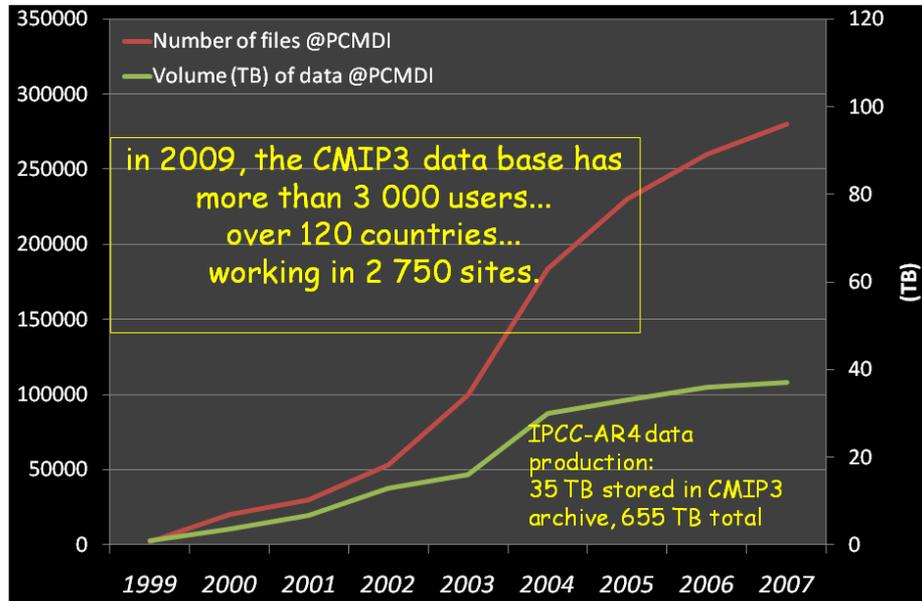
Example: CMIP5 long-term suite of experiments



source: B. Lawrence and K.Taylor

CMIP5, l'inflation !

- Des archives multi-sites (Fédération ESG)
- Des volumes de données en expansion



D'après G. Strand, GO-ESSP Meeting oct. 2009

CMIP5 (2011):
Requested output : **2PB !**
Replicated data : 1 PB

~ 220 TB decadal
~ 540 TB long term
~ 220 TB atmos-only

~100 TB of 3hourly atmos data!
~215 TB of ocean 3d monthly data!
~250 TB for the cloud feedbacks!

- Des modèles de climat de plus en plus complets (complexes) (*chimie atmosphérique, nuages, biogéochimie marine, végétation & cycle du carbone,...*)
- Une grande diversité de modèles (*résolution, dynamique, paramétrisations physiques, ...*)
- ...qui vont utiliser des forçages, des CI différents (« recommandations » CMIP5)

Objectifs METAFOR

Construire un modèle d'information standard capable de décrire : (WP2)

- les données issues des simulations climatiques;
- les simulations ayant produit ces données;
- les modèles utilisés pour ces simulations.

CIM :
« Common
Information
Model »

Promouvoir ce standard auprès de la communauté climatique

Développer des outils de création d'instances CIM (WP6)

- Questionnaire CMIP5 : collecte des métadonnées sur les modèles et simulations)
- Interface d'instanciation du CIM : Geonetwork

Ce sont ces « instances CIM » qui constitueront l'archive de métadonnées CMIP5

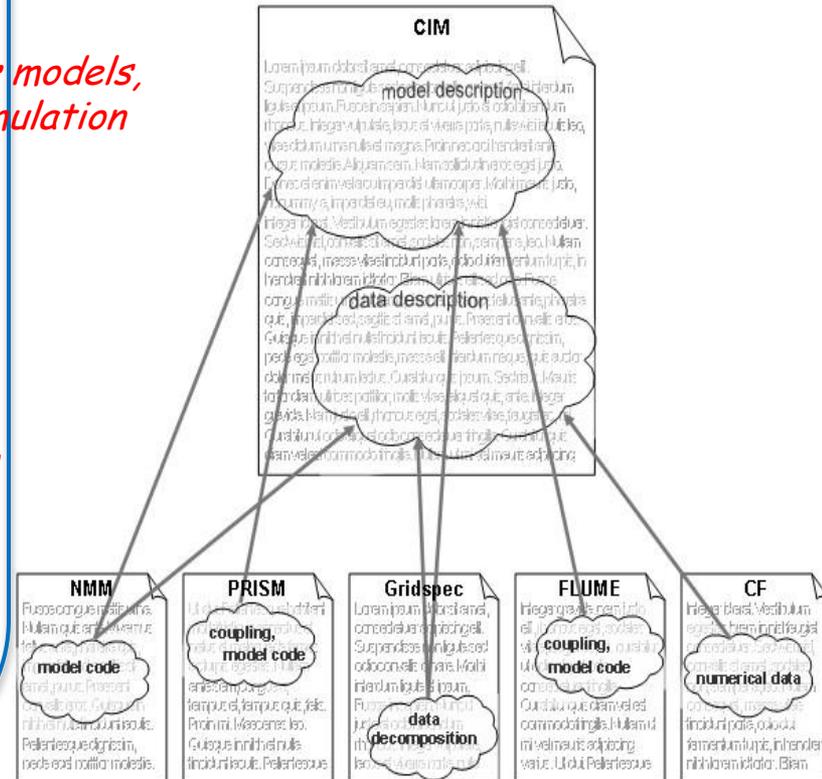
Développer des services d'accès et outils de manipulation des métadonnées (WP4/WP5)

- services sécurité/authentification, outils d'ingestion et validation
- opérations de visualisation, recherche et requête sur la base

Ces services et outils sont déployés sur le portail METAFOR

Les « standards » existant

- **CF-NetCDF** (Climate Forecasting conventions for NetCDF, PCMDI/NCAR-UCAR) *standard names for variables*
- **NumSim** (Numerical Simulation Discovery Metadata, BADC/NCAS) *ontology for models, code and simulation*
- **NMM** (Numerical Model Source Metadata, BADC/NCAS) *ontology for models, code and simulation*
- **CCMVal** (Chemistry-Climate Model Validation Activity, NCAS) *vocab for atmos chemisrty model*
- **FLUME** (Met-Office) *vocab for atmos chemisrty model*
- **ESG** (Earth System Grid) *ontology for model code*
- **CURATOR** (NOAA/NCAR/GFDL) *Model run configuration*
- **CERA2** (MPI) *data management*
- **BFG** (Univ. Of Manchester) *standard coupling interfaces*
- **ESMF** (Earth System Modeling Framework, NASA/NOAA) *standard coupling interfaces*
- **PRISM/OASIS4** (Max Planck Institute/CERFACS) *standard coupling interfaces*
- **GridSpec** *ontology for grids*
- **CMOR** (Climate Model Ouput Rewriter, NOAA/BADC/PCMDI) *CMIPs output data*
- *CMIPs output data*

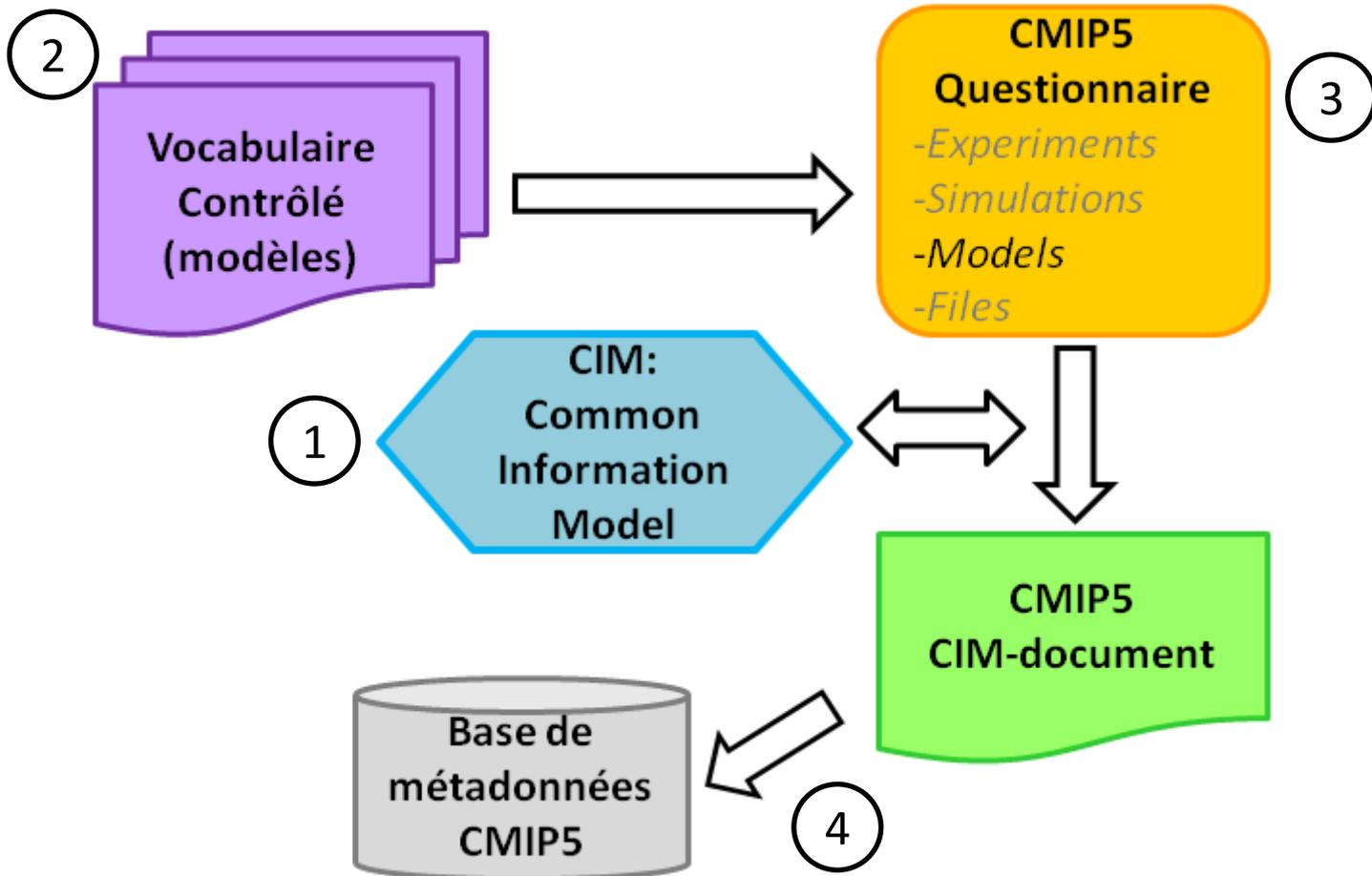


Pas de standard unifié
(fragmentés, incomplets, institution-dependant)

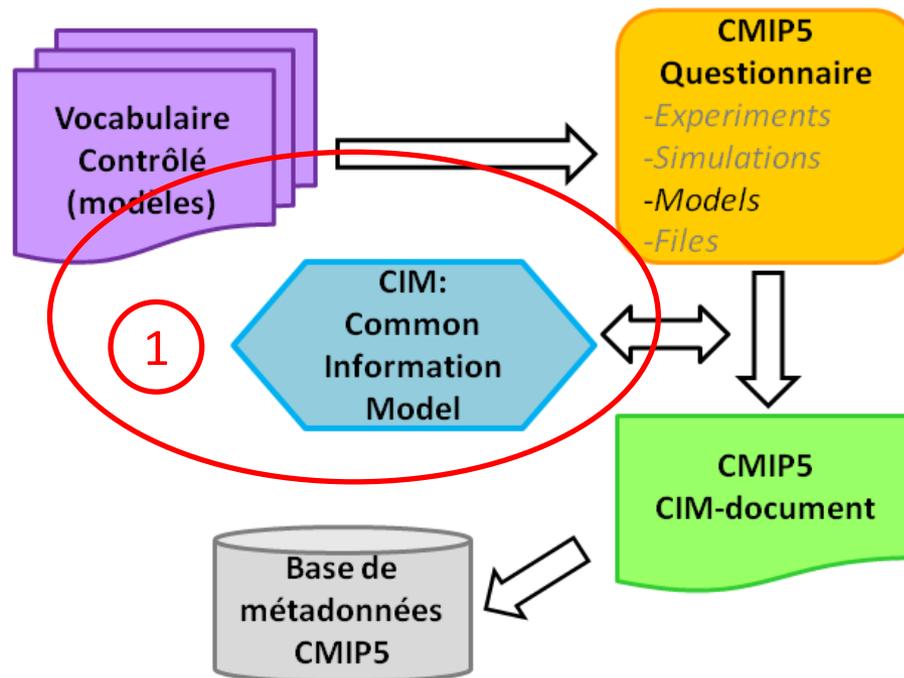


Rôle de METAFOR: coordonner l'unification (assurer les correspondances, éviter les duplications, **créer de nouvelles métadonnées**)

Déroulement de la présentation



Le CIM



Le CIM conceptuel (UML)

GridPackage

Description des grilles des modèles ou des données

DataPackage

Description des données produites et utilisées (forçages, CI, ...)

ActivityPackage

Description des expériences et des simulations

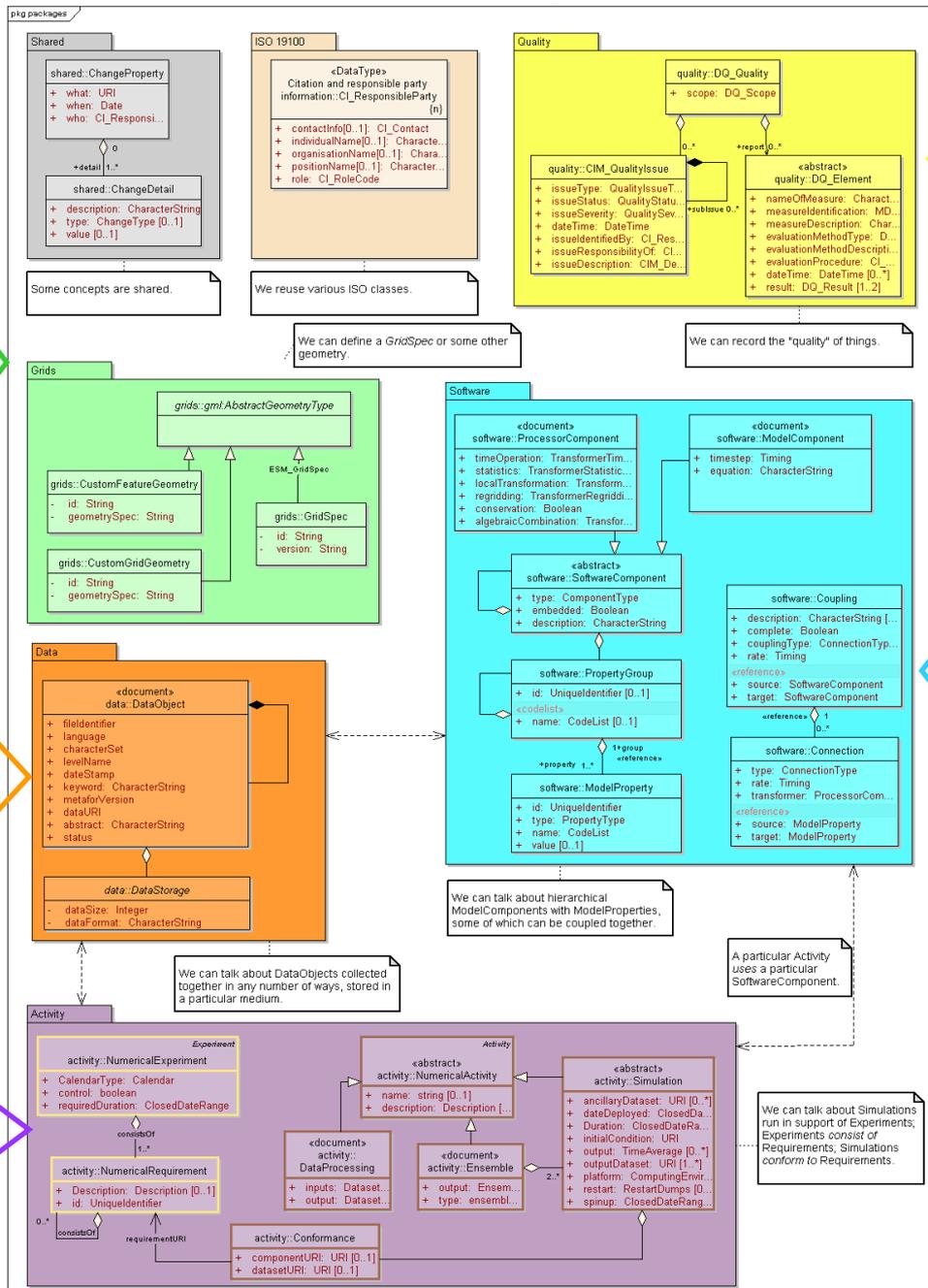
QualityPackage

Description de la qualité des données

SoftwarePackage

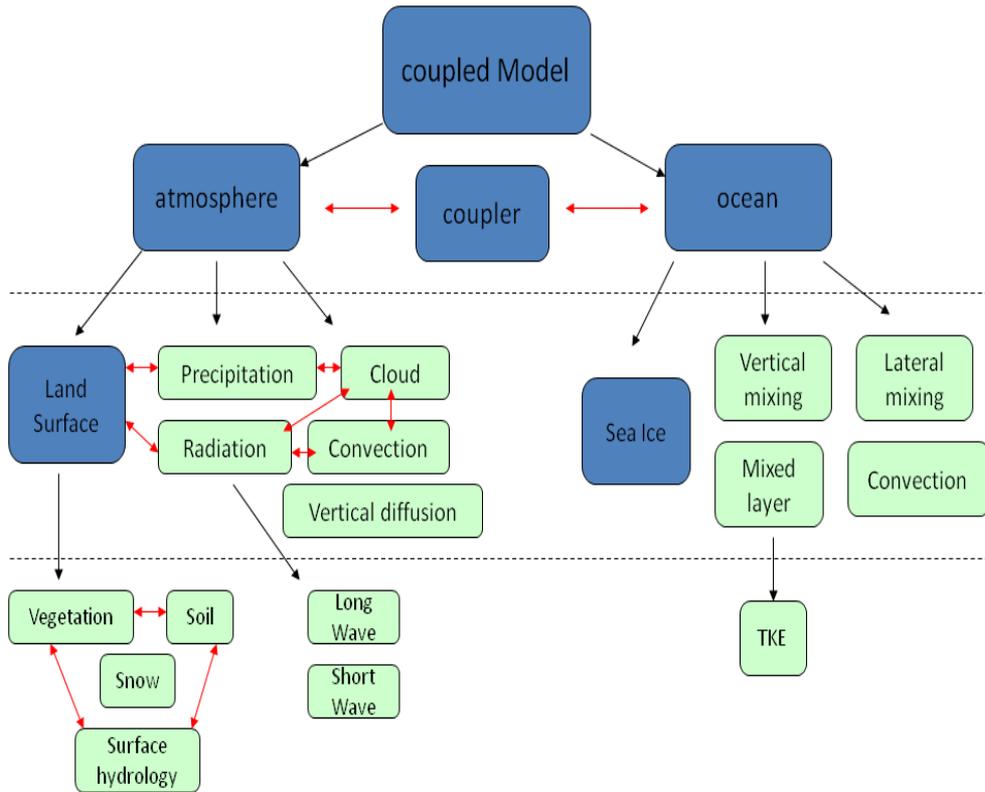
Description de toutes les composantes modèles (sous-composantes, coupleur, ...)

Représenter l'activité de modélisation climatique comme un ensemble de concepts structurés et liés entre-eux

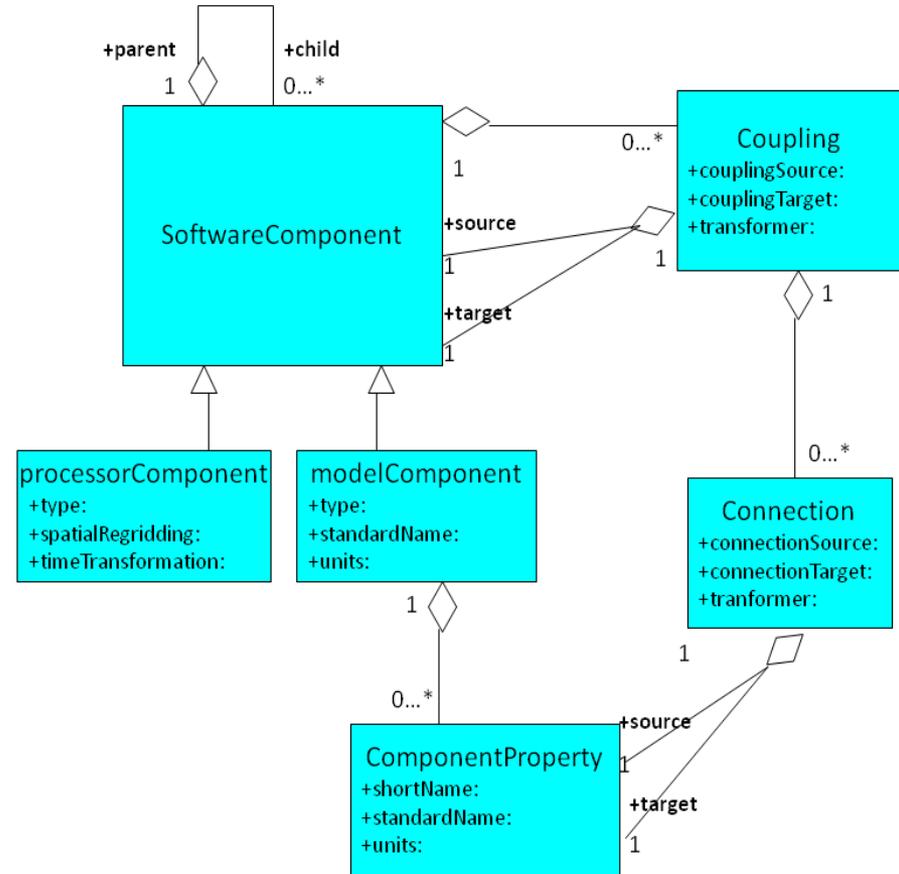


Description des modèles : Software package du CIM

Modèle couplé CMIP5



CIM Software Pkg



Description des simulations : Activity package du CIM

Simulation CMIP5

Experiment:

“1.1 decadalYYYY: ensembles of 10-years hindcasts and predictions”

source: CMIP5 Experiment Design, Karl Taylor

Requirements:

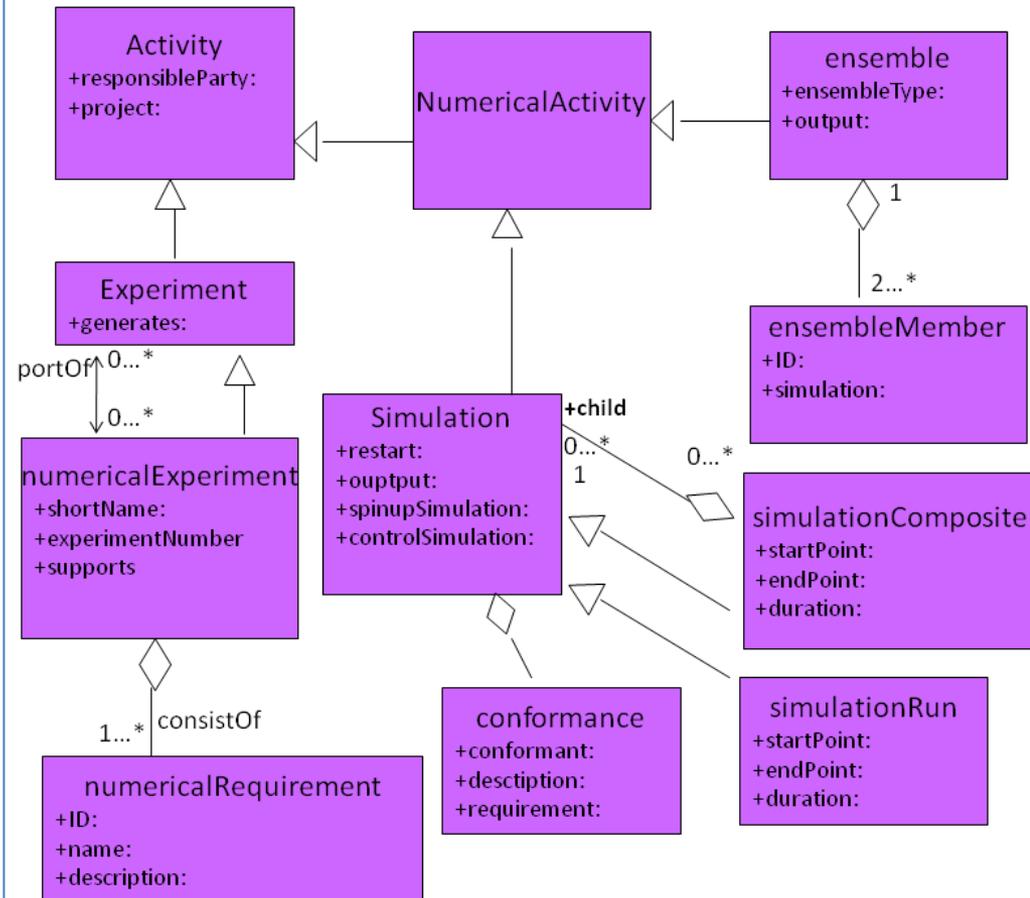
- Ocean initialization in some way representative of the observed anomalies
- Initialization dates : 1960 – 2005 every 5 years
- The atmospheric composition should be prescribed as in the historical run and the RCP4.5 scenario

Simulation décennale d'ensemble "i1"

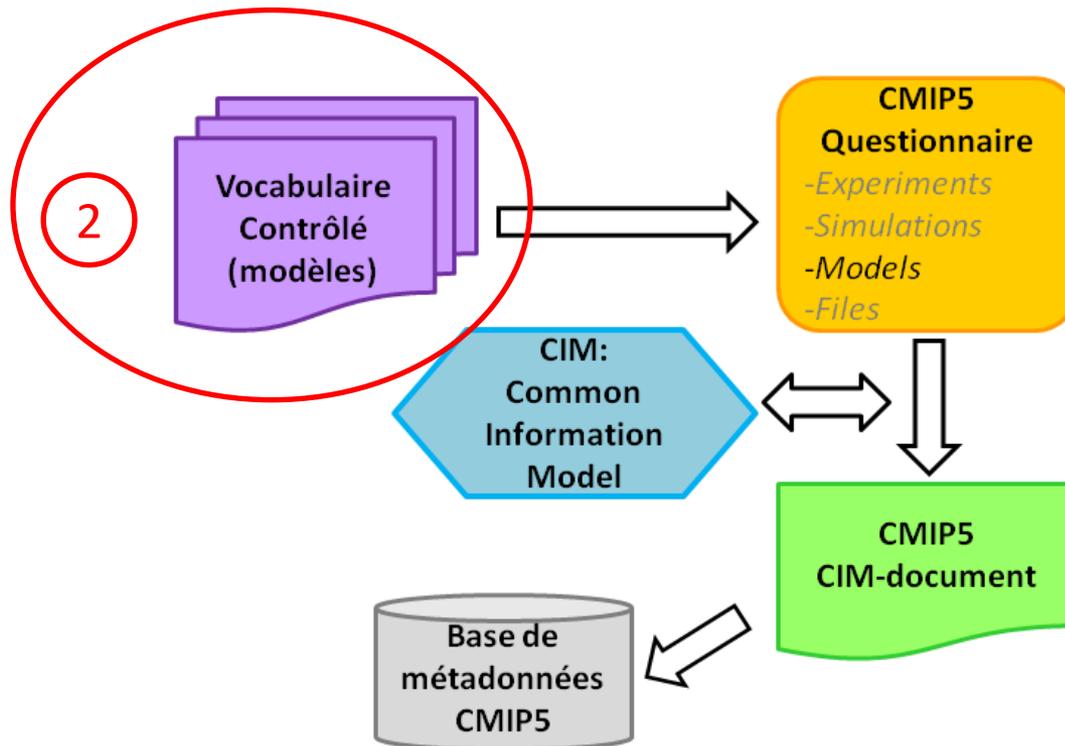
<r[1-9]i1p1> du CERFACS:

- initialisée par un run historique nudgé par rappel à la réanalyse pendant 3 ans (HISTNUD15).
- La simulation de spinup est SPEK2 et le run de controle PICTL.
- Les états atmosphériques qui distinguent les membres sont donnés par la simulation historique HIST.

CIM Activity Pkg



Vocabulaire Contrôlé



Le vocabulaire contrôlé (CV)

Spécialise et contraint le CIM pour une application particulière :

« les simulations climatiques dans le contexte de CMIP5 »

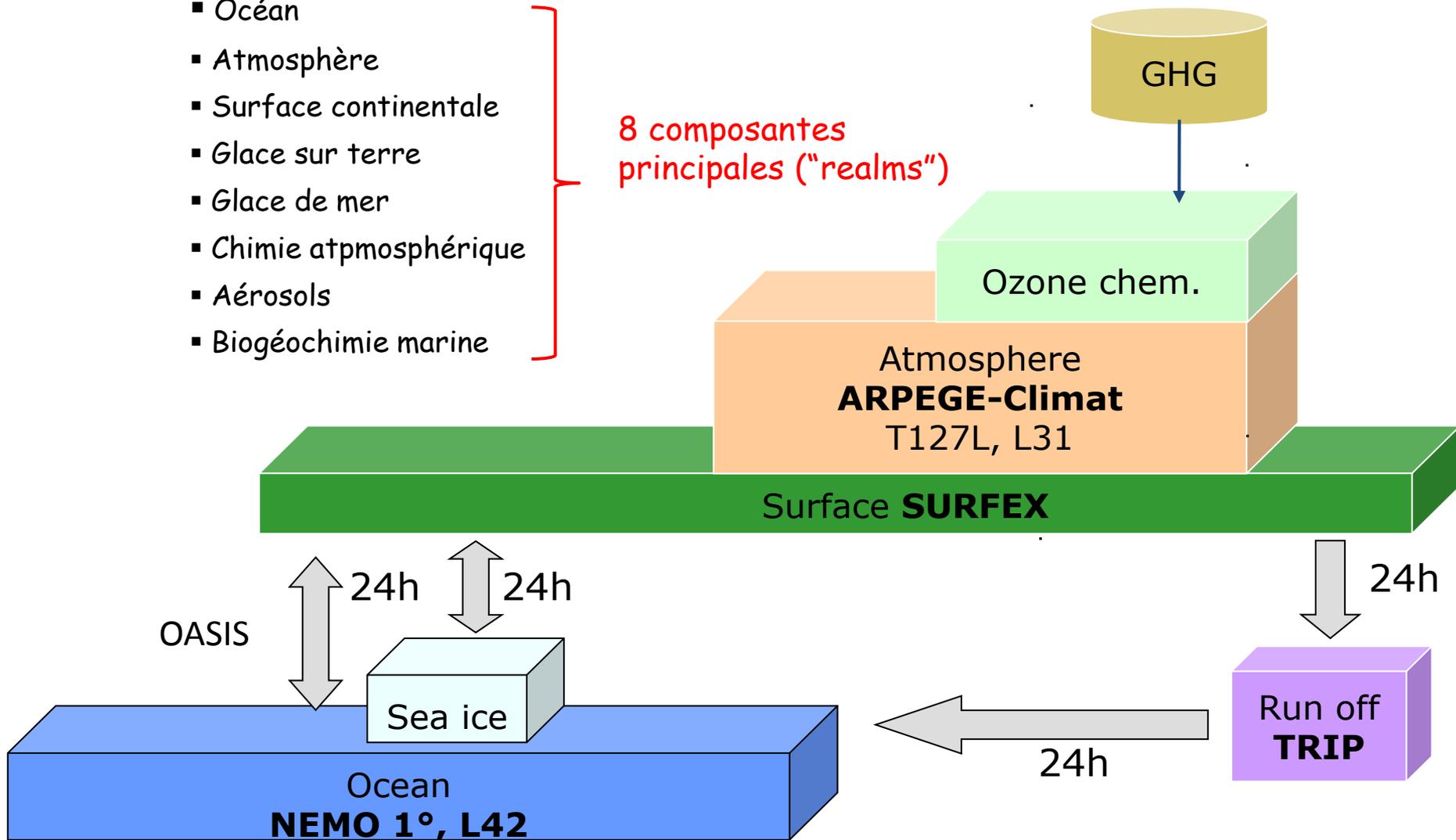
Type de CV pour CMIP5	Exemple	Défini par
Model names (institute)	HadGEN2-ES(MOHC), ESM2G(GFDL), CCSM4(NCAR), CESM(NCAR), CNRM-CM5(CNRM/CERFACS)	CMIP5/PCMDI
Experiment names and requirements	3.2 historical, 3.3-E amip, 3.4 midHolocene, 1.1-E decadal1965, 1.4 volcIn2010	CMIP5 (Experiment Design)
Model component names	atmos, ocean, land, landice / atmosRadiation, oceanAdvection, riverRouting, LandSurfaceEnergyBalance	CMIP5 DRS (realm only) + METAFOR
Model component content	NumberOfGroundWaterLayer, WaterStorageMethod	METAFOR +ESG ontology
Output variables	sea_surface_temperature, sea_water_salinity, rate_of_melt_at_upper_surface_of_sea_ice	CMIP5 (CF standard names)
Transformations	Linear, near-neighbour, conservative-first-order / TimeAccumulation, TimeInterpolation	OASIS/BFG + METAFOR
Grid characteristics	Terrain-following, mass-based / latitude-longitude, regular gaussian, displaced pole	GridSpec + METAFOR + CMOR2
Output files	project_id, modeling_realm, forcing, initialization_method	CMOR2

Les Modèles de Climat à documenter: ex. CNRM-CM5

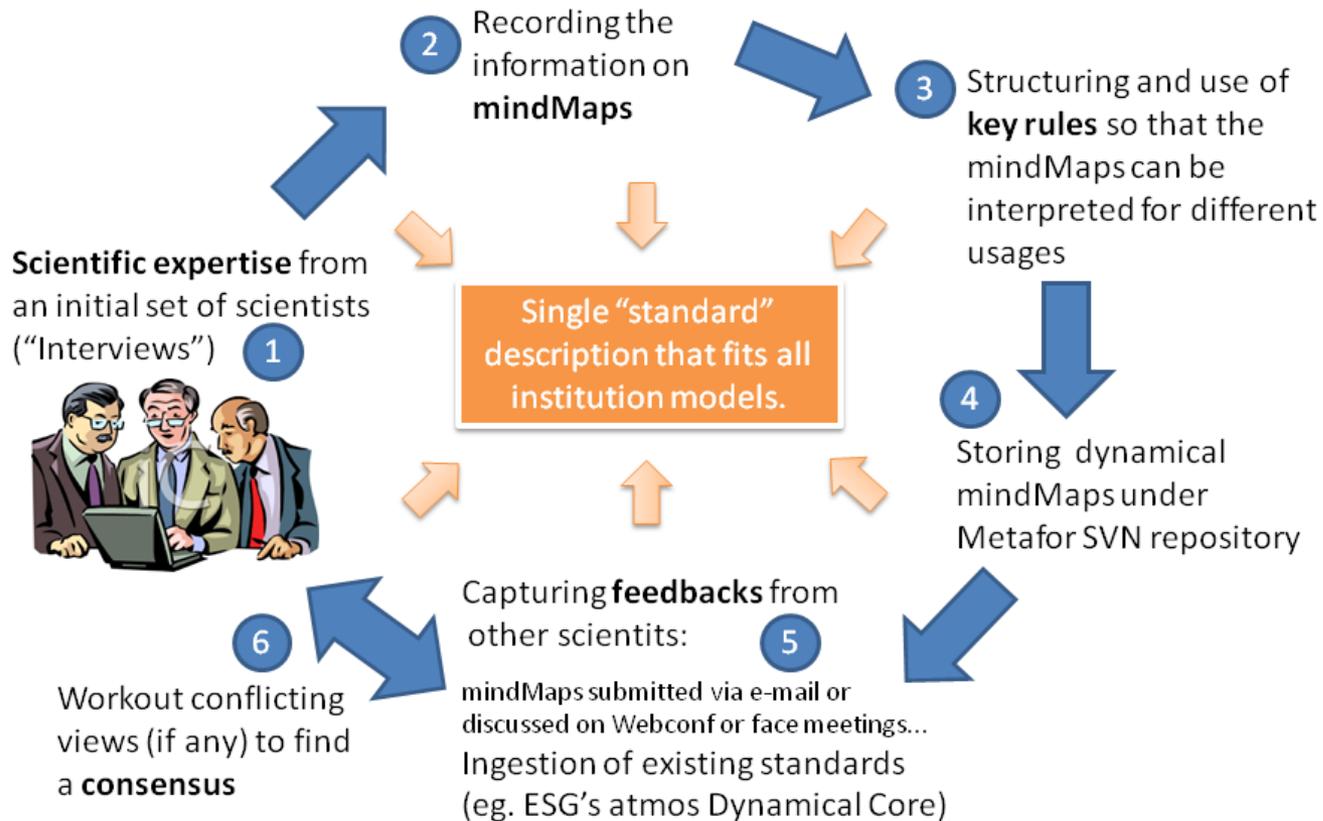
Modèles couplés (ESM ou GCM) :

- Océan
- Atmosphère
- Surface continentale
- Glace sur terre
- Glace de mer
- Chimie atmosphérique
- Aérosols
- Biogéochimie marine

8 composantes principales ("realms")



Determination du CV "scientifique"



- > 35 contributeurs scientifiques, experts en modélisation du Climat
- 13 centres de recherche, 6 pays

Contributeurs

Atmosphere

Michel Déqué CNRM, France
Jan Polcher, LMD-IPSL, France
Fredéric Hourdin , LMD-IPSL, France
Bruce Wyman, GFDL, USA
Bryan Lawrence, NCAS-BADC, UK (*DynCore*)
Masa Kageyama, LSCE-IPSL, France (*Paleo*)
Olivier Marti, LSCE-IPSL, France (*Paleo*)
Myriam Khodry, IPSL, France (*Paleo*)

Land surface

Bertrand Decharme CNRM, France
Jan Polcher, LMD-IPSL, France
Aaron Boone, CNRM, France (*Land snow*)
Anne-Laure Gibelin, CNRM, France (*C.cycle+Veg.*)
Sergey Malyshev, GFDL, USA
Balaji, GFDL, USA
Stefan Hagemann, MPI, Germany
Gerhard Krinner, LGGE, France (*Land snow*)
Masa Kageyama, LSCE-IPSL, France (*Paleo*)
Olivier Marti, LSCE-IPSL, France (*Paleo*)
Mark Elkington, MetOffice, UK (*Vegetation*)

Land ice

Jan Polcher, LMD-IPSL, France
Catherine Ritz, LGGE, France
Gerhard Krinner, LGGE, France
Masa Kageyama, LSCE-IPSL, France (*Paleo*)
Olivier Marti, LSCE-IPSL, France (*Paleo*)

Ocean biogeochemistry

Laurent Bopp, LSCE-IPSL, France

Ocean

Gurvan Madec, LOCEAN-IPSL, France
Eric Guilyardi, LOCEAN-IPSL, France
Stephen Griffies, GFDL, USA
Balaji, GFDL, USA
Masa Kageyama, LSCE-IPSL, France (*Paleo*)
Olivier Marti, LSCE-IPSL, France (*Paleo*)

Atmospheric chemistry

Vincent-Henri Peuch, CNRM, France
Luke Abrahams, UKCA, UK
Larry Horowitz, GFDL, USA
Balaji, GFDL, USA
Peter Van Velthoven, KNMI, Netherland
Peter Braesicke, UKCA, UK
Christoph Bruehl, UKCA, UK
Ross Slawitch, Univ. Maryland, UK
Graham Mann, Univ. of Leeds

Aerosols

[same as above + reviewed by...]

Vincent-Henri Peuch, CNRM, France
Graham Mann, Univ. of Leeds

Sea ice

David Salas Y. Melia, CNRM, France
Martin Vancoppenolle, UCL-LLN, Belgium
Hugues Goosse , UCL-LLN, Belgium
Thierry Fichefet , UCL-LLN, Belgium

+ Lawrence Buja/Gary Strand, NCAR, USA

CNRM

IPSL

LGGE

UKCA

Univ. Leeds

Univ. Maryland

BADC

MetOffice

GFDL

NCAR

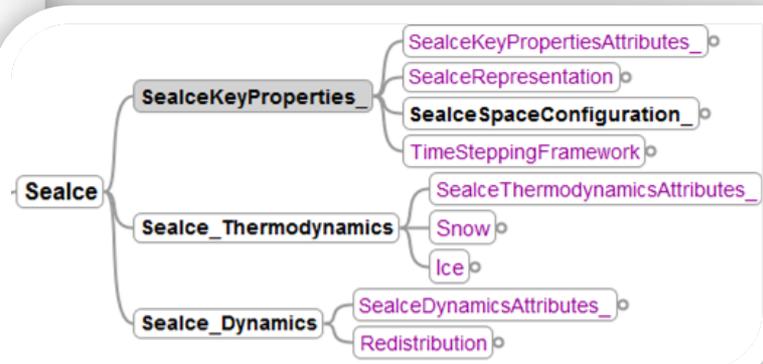
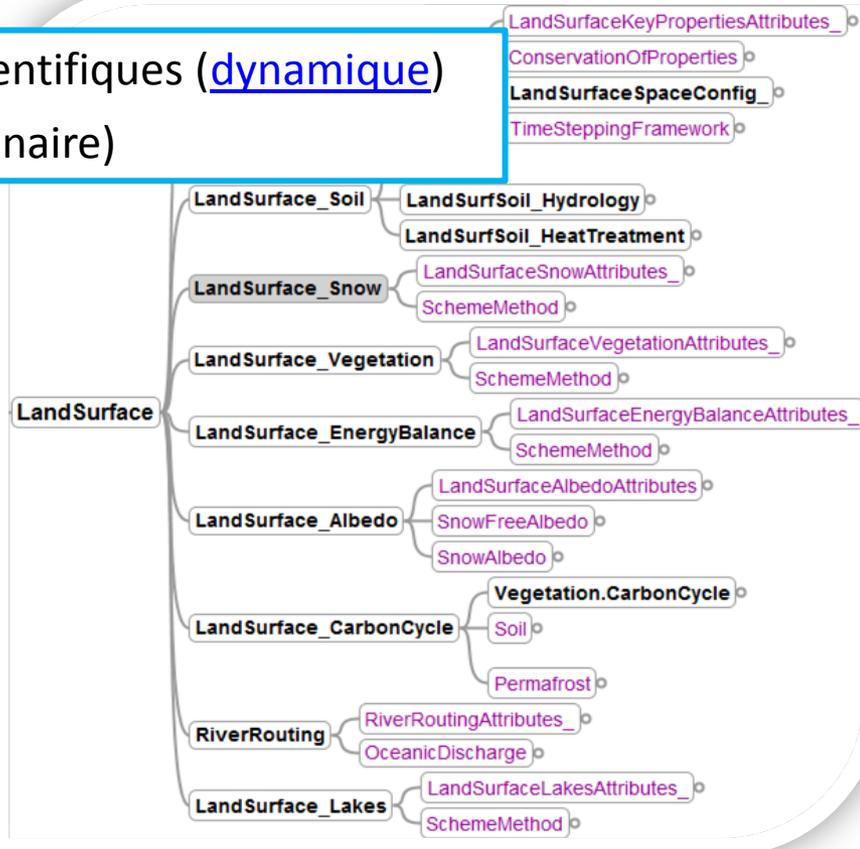
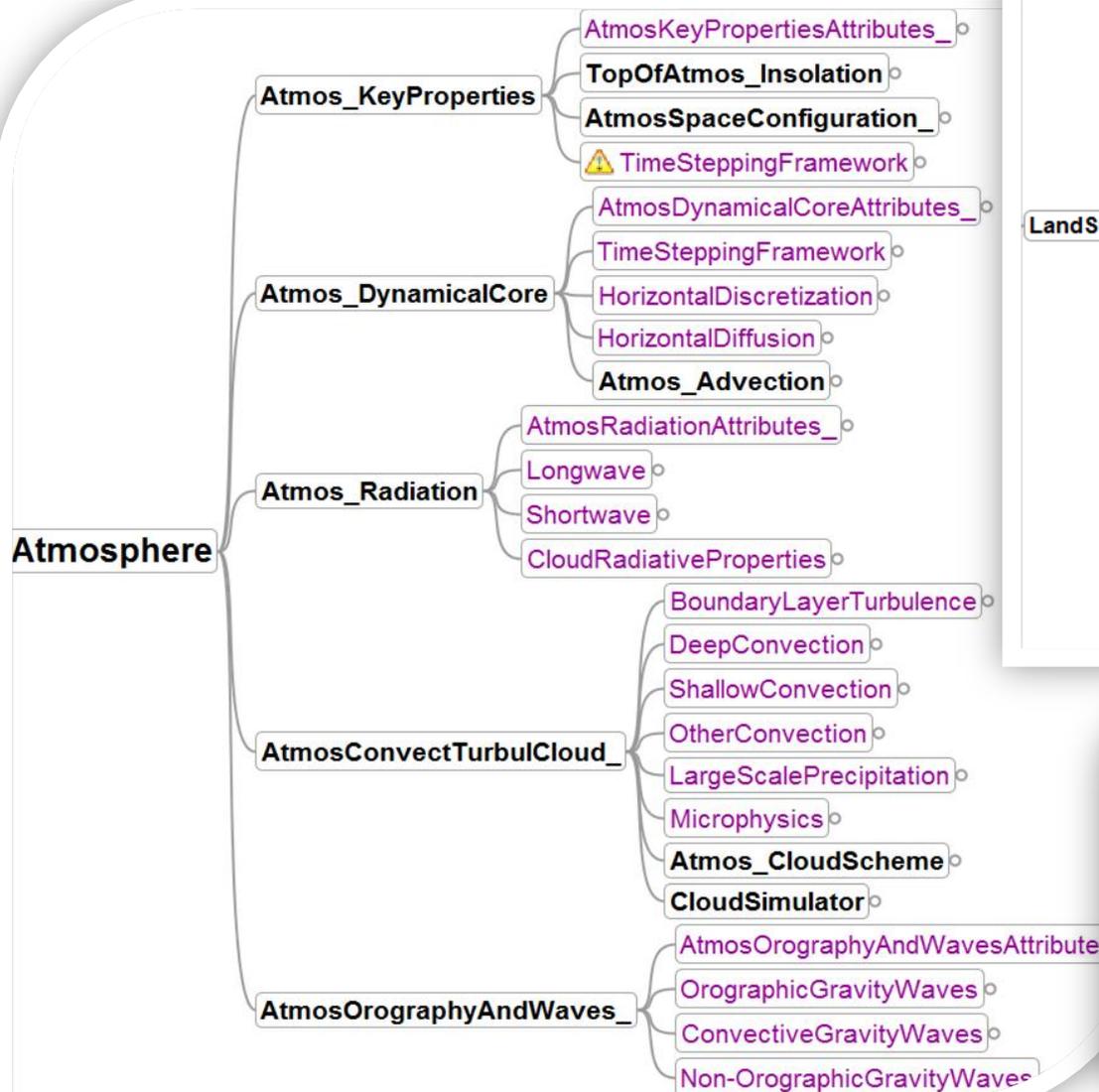
MPI

LLN

KNMI

CV consigné au format « mindmap »

- bon support d'interaction/dialogue avec les scientifiques (dynamique)
- interprétable par des outils Python (-> questionnaire)



Organisation du CV

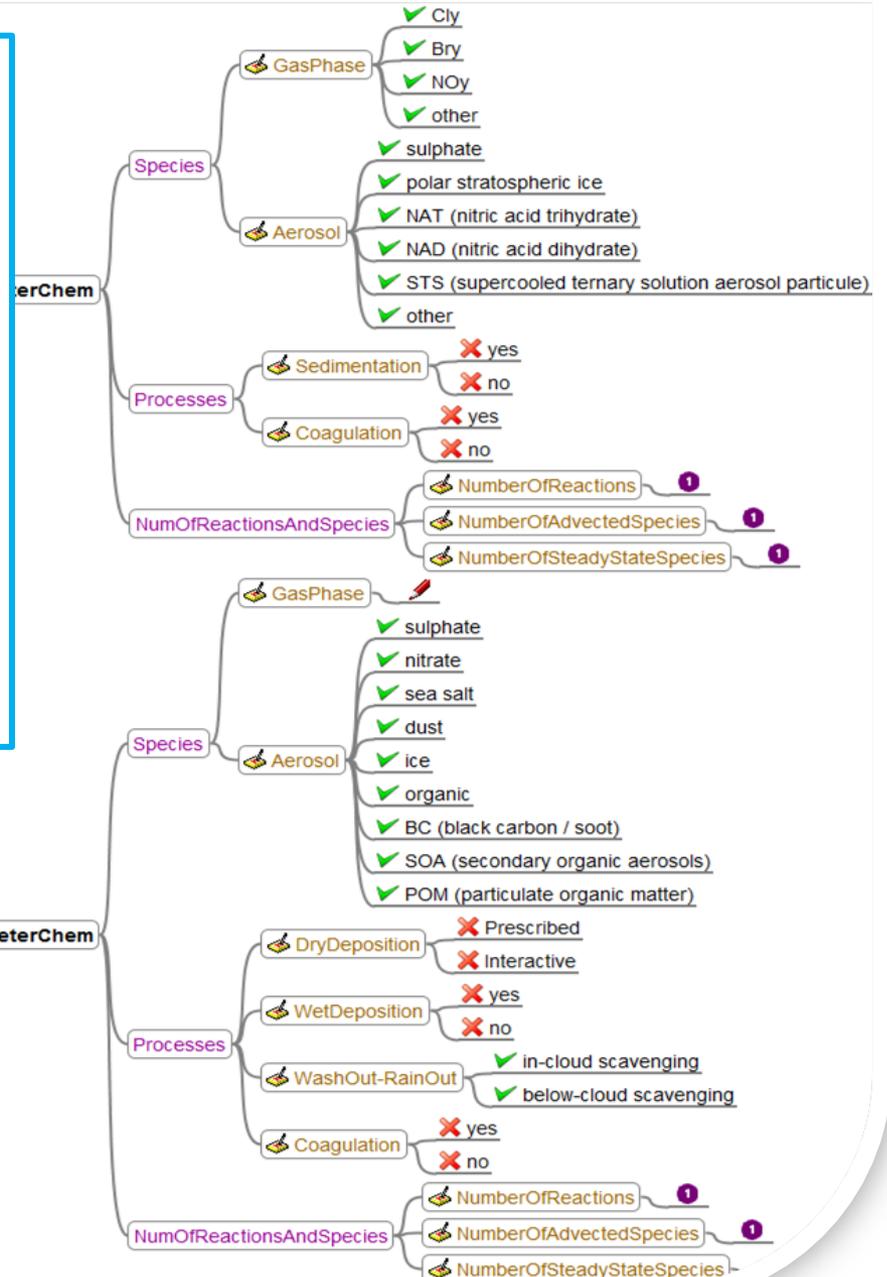
Structure

L'agencement du CV est dictée par la logique scientifique des modèles mais tient aussi compte de contraintes techniques du Questionnaire (Django)

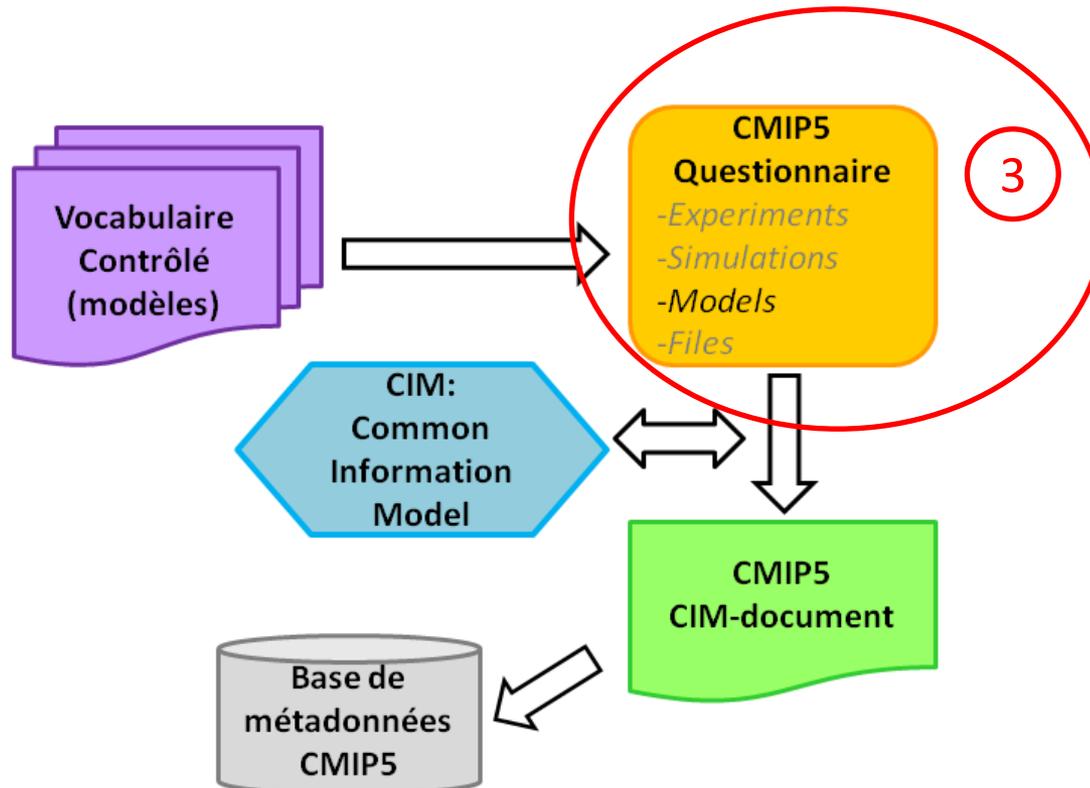
Règles de codages

- Une composante "realm" est constituée d'une **composante** / schéma de paramétrisation
- Elle-même composée de **groupes de paramètres** pour rassembler des **paramètres** caractérisant la même entité.
- Des listes de valeurs possibles sont associées aux **paramètres**

Tropospheric.HeterChem

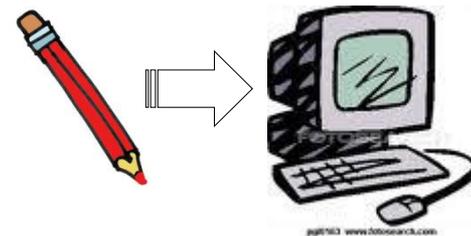


Questionnaire CMIP5



Le questionnaire CMIP5

- Metafor mandaté par le WCRP/CMIP panel
- questionnaire web
- code Django/Python (BADC/NCAS)
- phase de beta-test
- première version officielle 1.0 disponible depuis le 04/11



CMIP3
questionnaire

CMIP5
questionnaire

Les renseignements demandés :

- ✓ Le contenu scientifique des composantes-modèle : schémas de paramétrisations, schémas numériques (y compris les grilles)
- ✓ Les données externes et les détails du couplage (conditions initiales, forçages, conditions aux limites, ...)
- ✓ Chaque simulation (chacun de ses membres)
- ✓ La façon dont chaque simulation respecte les spécifications de l'expérience CMIP5 à laquelle elle se rattache (« conformance »)

Questionnaire hébergé au BADC :

<http://q.cmip5.ceda.ac.uk/>



Questionnaire: Page d'accueil

Summary Experiments Model Grid Simulation Files References Parties Help

The screenshot shows a web interface for a questionnaire. At the top is a navigation bar with tabs: Summary, Experiments, Model:GCM 1, Grid:Grid1, Simulation (highlighted in yellow), Files, References, Parties, Help, and About. Below the navigation bar is a main content area with a yellow background. The content is organized into several sections:

- Summary: Dummy Centre used to hold examples (Example)**: A heading for the main content.
- Introduction**: A text block explaining the purpose of the questionnaire and the relationship between models, platforms, and simulations.
- Computing platforms associated with Example**: A table with one row containing 'machine1' and an 'edit' button. Below it is an 'Add a new Platform' button. A green callout box labeled 'Platform 1st' points to this section.
- Grids associated with Example**: A table with three rows: 'Grid1', 'Grid1cp', and 'Grid Template dup'. Each row has 'Edit' and 'Copy' buttons. Below it is an 'Add a new grid' button. A purple callout box labeled 'Grids 2nd' points to this section.
- Models associated with Example**: A table with two rows: 'GCM 1' and 'GCM 2'. Each row has 'Edit' and 'Copy' buttons and a 'Status' column showing '0.0% valid'. Below it is an 'Add a new model' button and a note: 'It can take some time (minutes) to create a new model or copy an existing one...'. A red callout box labeled 'Model 3rd' points to this section.
- Files, References and Parties**: A section with three lines of text: 'There are 0 references associated with Example', 'There are 2 files associated with Example', and 'There are 1 people and institutions associated with Example'. A grey callout box labeled 'Inputs 4th' points to this section.
- Simulations associated with Example**: A section with one line of text: 'Manage Simulations (add new)'. A blue callout box labeled 'Simulations 5th' points to this section.

An orange callout box at the top right contains the text: 'Yellow tab text indicates that this element still needs to be created on the summary page. Only then can the tab be used for navigation'. An arrow points from this box to the 'Simulation' tab in the navigation bar.

Screenshot of an institutes Summary (or hub) page

Questionnaire: Page « Modèle »

CMIP5 Model Metadata...

Model Component OceanAdvection

Components Pages

Components hierarchy

Model tree

Available Models

CERFACS_VMI

- Aerosols
- Atmosphere
- AtmosphericChemistry
- LandIce
- LandSurface
- OceanBiogeoChemistry
- Ocean
 - OceanKeyProperties
 - OceanAdvection**
 - OceanLateralPhysics
 - OceanVerticalPhysics
 - OceanUpAndLowBoundaries
 - OceanBoundaryForcing
- SeaIce

Component OceanAdvection

Please add details of any other relevant subcomponents of this component

Parameters

Attributes

The questionnaire is organized into pages which further describe this component.

Inputs Needed

References

The buttons in this box navigate to pages for this model.

Resolve Inputs

View

Validate

Export XML

Short Name: OceanAdvection (type: OceanAdvection)

Implemented: Untick the box if there is no representation of OceanAdvection in your model.

Long Name:

Responsible Parties (Use the parties tab to add more choices here):

Contact:

Lead Author:

Funder:

Component Attributes

In this section enter parameters and attributes associated with this component.

General Attributes

Enter an additional component attribute or parameter if you wish (you can enter multiple parameters by entering one at a time):

Name

Value

Momentum

SchemeType: Choose one of:

Flux form
Vector form

SchemeName: Enter string value::

Enter an additional component attribute or parameter if you wish (you can enter multiple parameters by entering one at a time):

Name

Value

LateralTracers

SchemeType: Choose one of:

MonotonicFluxLimiter: Choose one of:

yes
no

Enter an additional component attribute or parameter if you wish (you can enter multiple parameters by entering one at a time):

Name

Value

VerticalTracers

SchemeType: Choose one of:

MonotonicFluxLimiter: Choose one of:

yes
no

Enter an additional component attribute or parameter if you wish (you can enter multiple parameters by entering one at a time):

Name

Value

Parameters groups

Frames

Parameters potential values

Attributes potential values

Centred 2nd order

Centred 4th order

Total Variance Dissipation (TVD)

etc...

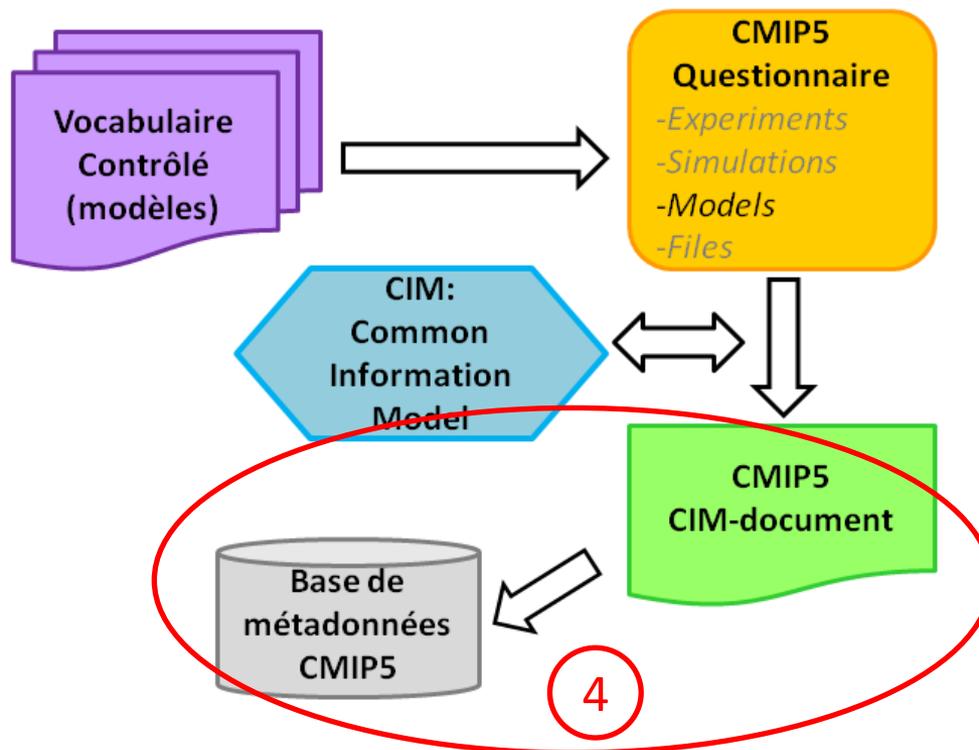
Centred 2nd order

Centred 4th order

Total Variance Dissipation (TVD)

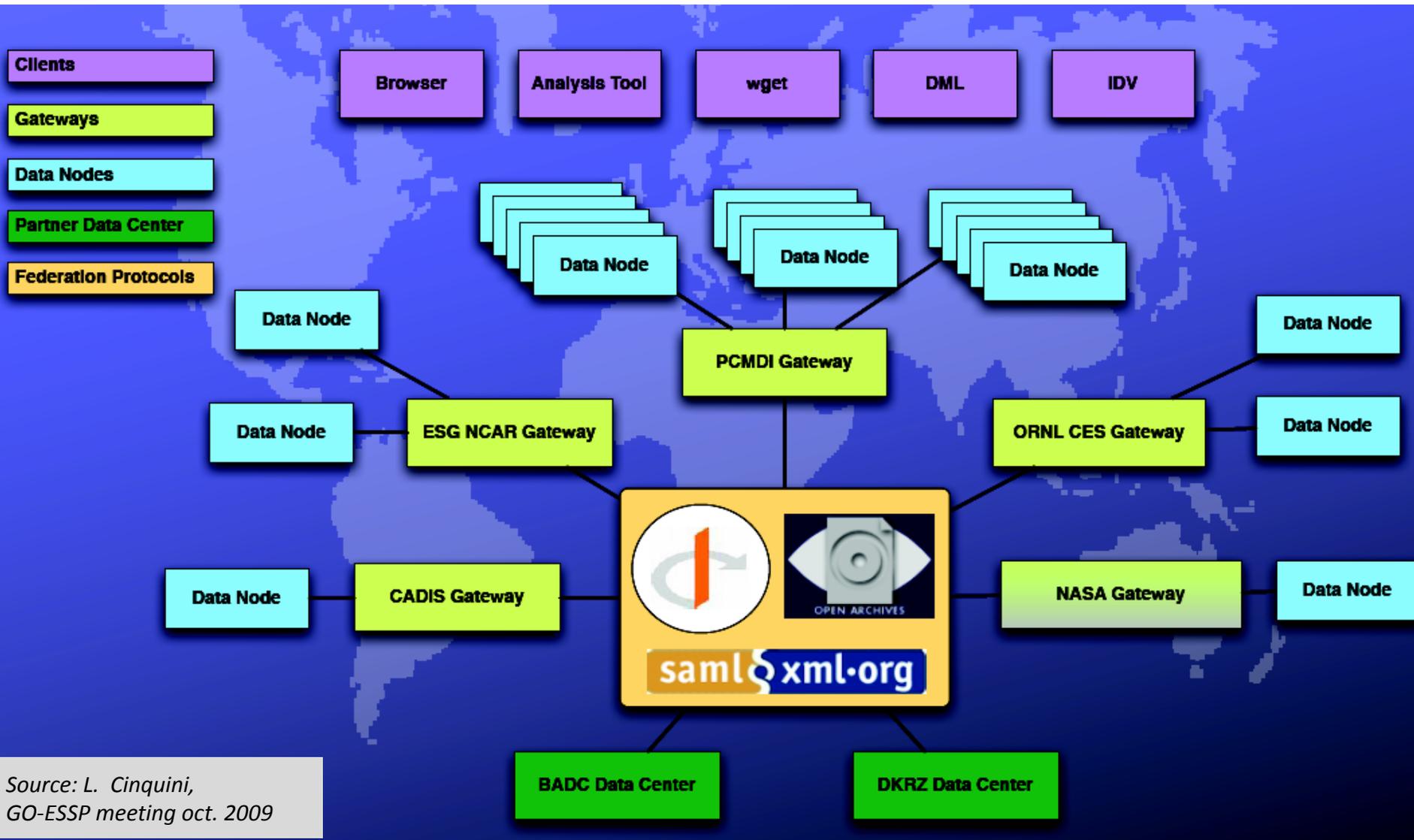
etc...

Archivage et accès aux métadonnées



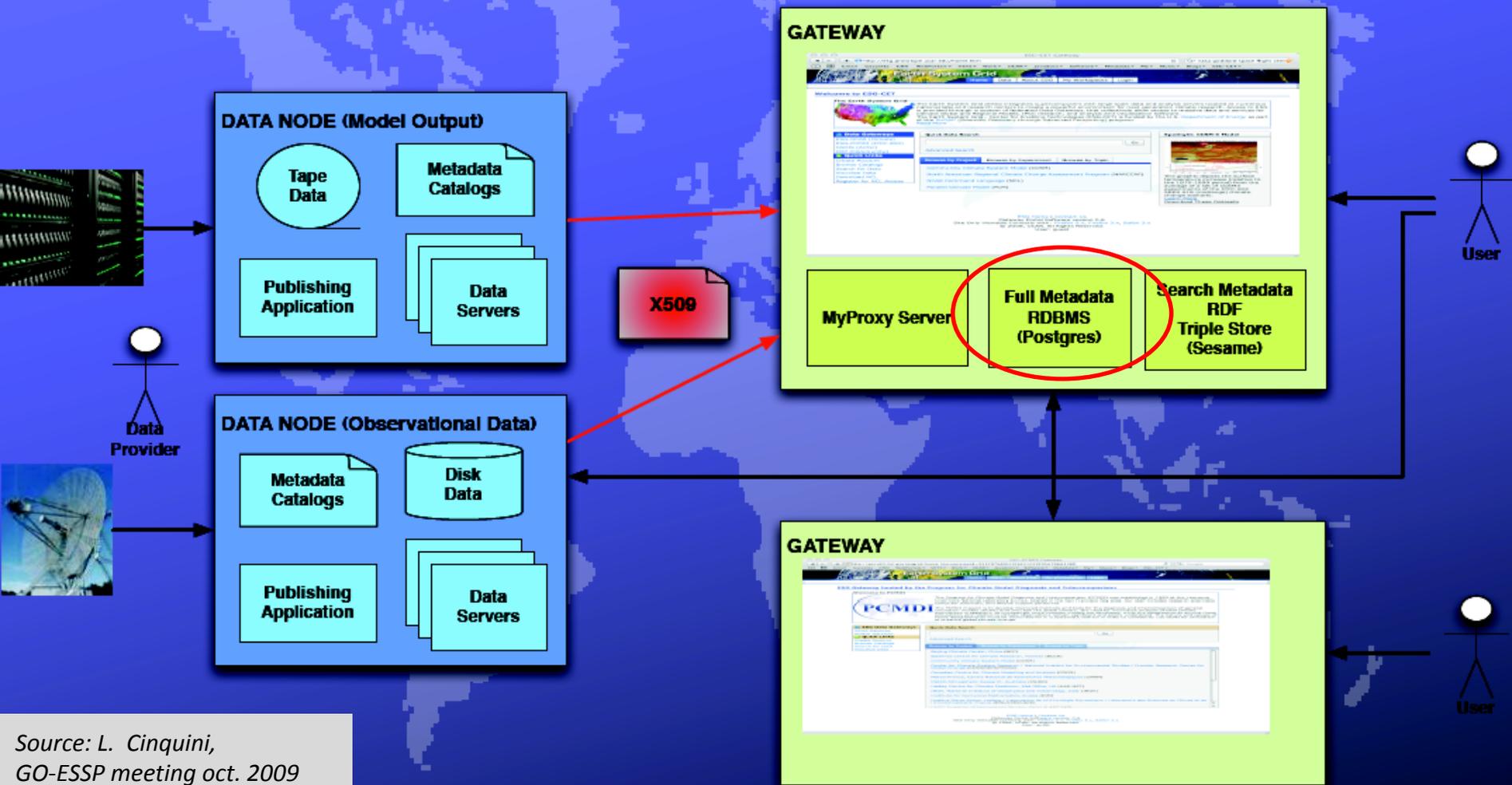
Données CMIP5 distribuées - fédération ESG

“Distributed, Federated, Integrated, Extensible”



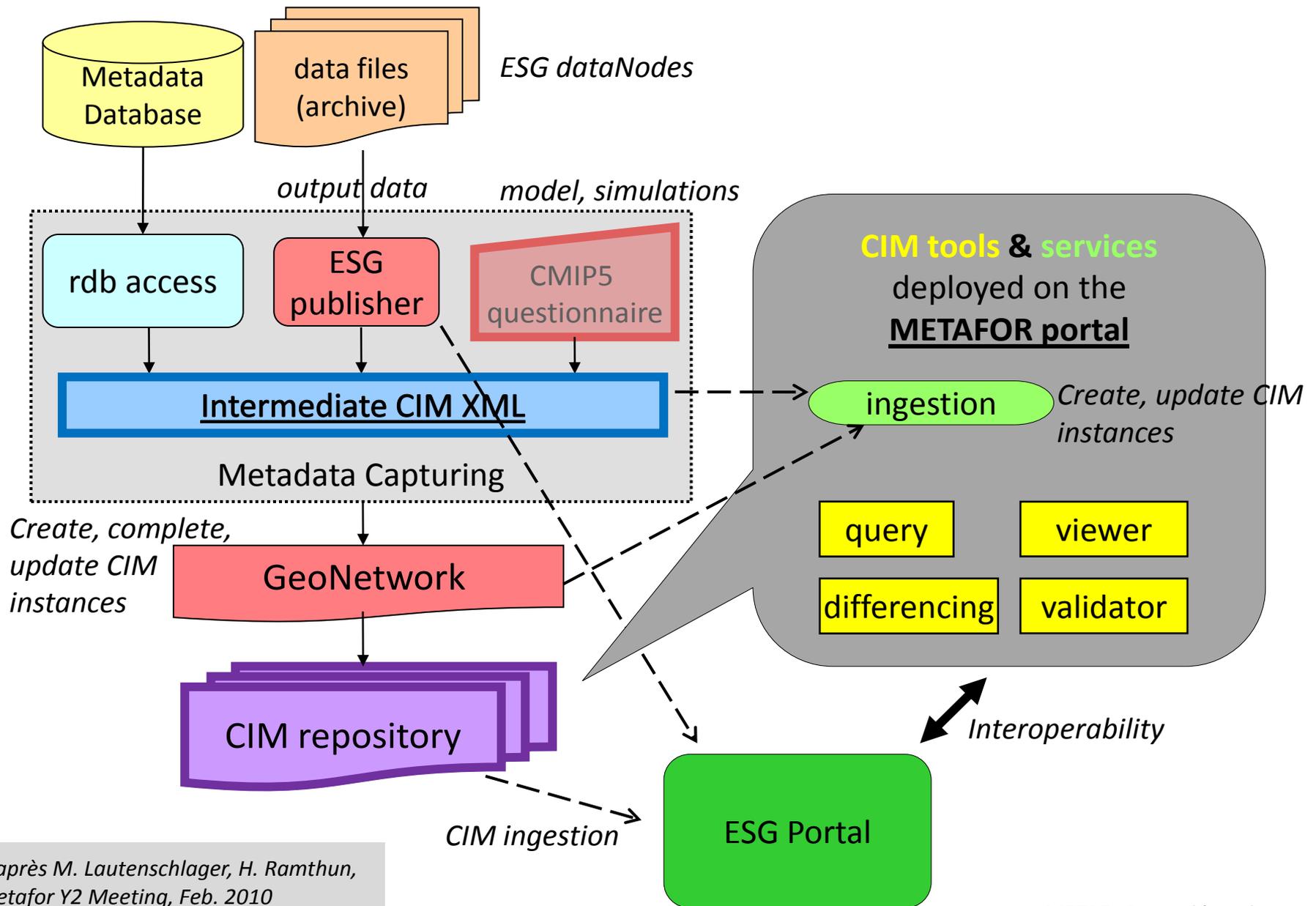
Source: L. Cinquini,
GO-ESSP meeting oct. 2009

Accès aux métadonnées au sein d'ESG-CET



Source: L. Cinquini,
GO-ESSP meeting oct. 2009

Alimentation de la base de métadonnées CMIP5



D'après M. Lautenschlager, H. Ramthun,
Metafor Y2 Meeting, Feb. 2010

Lien métadonnées - outputs

CMIP5/output/IPSL/IPSLCM5A/historical/da/atmos/pr/r1i1p1

```
// global attributes:
      :institution = "IPSL (Institut Pierre Simon Laplace,
Paris, France)";
      :*_institute_id*_ = "IPSL";
      :*_experiment_id*_ = "historical";
      :source = "IPSLCM5A (2010) : ...";
      :*_model_id*_ = "IPSLCM5A";
      :forcing = "Nat, Ant, GHG, SA, Oz, LU, SS, Ds, BC, MD
OC, AA";
      :parent_experiment_id = "piControl";
      :branch_time = 1000.;
      :contact = "Sebastien Denvil,
sebastien.denvil@ipsl.jussieu.fr";
      :history = "data generated 2010-05-04T10:15:36Z CMOR
rewrote data to comply with CF standards andCMIP5 requirements." ;
      :comment = "This 20th century simulation was initiated
from a preindustrial control simulation when equilibrium was reached ...";
      :references = "Model documentation and further
reference available here ...";
      :*_initialization_method*_ = 1;
      :*_physics_version*_ = 1;
      :tracking_id = "59204209-dde4-49be-ab96-676c4fba9ee7";
      :product = "output";
      :experiment = "Historical";
      :frequency = "da";
      :creation_date = "2010-05-04T10:15:36Z";
      :Conventions = "CF-1.4";
      :project_id = "CMIP5";
      :table_id = "Table da (11 March 2010)
6e0a7e19df94506d06091b43082677dc";
      :title = "IPSLCM5A model output prepared for CMIP5
Historical";
      :parent_experiment = "pre-industrial control";
      :modeling_realm = "atmos";
      :*_realization*_ = 1;
      :cmor_version = "2.0.0";
```

Le questionnaire décrit
cette composante
"atmos" du modèle
"IPSLCM5A" sans
ambigüité

Le questionnaire demande
de spécifier les même
attributs identifiant la
simulation "r1i1p1"
(membre dans un
ensemble) réalisée dans le
contexte d'une expérience
"historical"

Conclusions

- Nous disposons d'un **modèle d'information (CIM)** pour la modélisation climatique, **complet et générique**
 - le CIM a atteint une bonne maturité mais est amené à évoluer au fil de l'extension de son champ d'application (en cours: données du projet ENSEMBLES, coupleurs BFG et OASIS4, downscaling statistique, etc...)
-
- Un travail collaboratif conséquent a été effectué **avec le projet ESG-CURATOR** pour rendre compatibles les 2 ontologies et adapter au CIM leur outils d'ingestion et de manipulation des métadonnées
 - Ceci a nécessité des efforts de **traduction-équivalence** avec les projets «associés» (CURATOR, GridSpec) mais aussi «en interne» (CIM/questionnaire)
-
- METAFOR met pour la première fois au service de la communauté climat **un jeu de métadonnées unifié** permettant de décrire de manière extensive les sorties de simulations climatiques (identification + **protocole détaillé de production**)
 - METAFOR apporte un progrès significatif dans la **description des modèles**
-
- **Le questionnaire CMIP5 est opérationnel.** Les groupes de modélisation sont invités à s'atteler à la tâche "métadonnées", ... Le processus est long, la logique du questionnaire parfois complexe et demande des flots de renseignements, mais **la publication des métadonnées est un passage obligé... et qui s'avèrera utile !**

Metafor, le film !

METAFOR

Common METAdata FOR climate modelling digital repositories



[CIM cartoon](#)