C O M P L E X
Knowledge Based Climate Mitigation Systems for a Low Carbon Economy

Model Repository

Report D6.1

Date	**14.10.2013**
Report Number	**D6.1**
VERSION NUMBER: Report Number: Main Authors:	**FINAL** D6.1 Alexey Voinov, Getachew Belete

DIFFUSION LEVEL – RIP	
PU	**PUBLIC**
RIP	**RESTRICTED INTERNAL AND PARTNERS**
RI	**RESTRICTED INTERNAL**
CO	**CONFIDENTIAL**

Coordinator: **Nick Winder, University of Newcastle**

INFORMATION ON THE DOCUMENT

Title	**Model Repository**
Author	Alexey Voinov, Getachew Belete
Co-authors	

DEVELOPMENT OF THE DOCUMENT

Date	Version	Prepared by	Institution	Approved by	Note
14.10.13	0.1		ITC-UTwente		
27.11.13	FINAL		ITC-UTwente		

INDEX

1. Introduction

As our impacts on the environment become more dramatic and the systems we analyze become more complex, there is a growing understanding that one model cannot be sufficient to represent all the details needed for decision making and planning (Argent, 2004; Gaber et al., 2008). There are also numerous legacy models that can be reused as building blocks for more complex systems, provided that they can be linked together in a meaningful way matching the variables, scales and resolutions. Currently there is a growing number of efforts to develop the standards and software tools that would provide for this kind of integration (Warner et al., 2008; Barthel et al., 2008; Argent et al., 2009). The US Environmental Protection Agency (EPA) has been developing the FRAMES (2009a, 2009b) (Framework for Risk Analysis in Multi-media Environmental Systems) system to manage the execution and data flow among multiple science modules. The Object Modeling System (OMS) is developed by the US Department of Agriculture (David et al., 2002; Kralisch et al., 2004; Ahuja et al., 2005). In contrast to FRAMES and some other systems, OMS requires rewriting modules in Java to be then inserted into the system library. The Open Modeling Interface and Environment (OpenMI, 2009) developed by a consortium of European universities and private companies, is a standard for model linkage in the water domain (Moore et al., 2005). The Common Component Architecture (CCA) is a product developed by the Department of Energy and Lawrence Livermore National Lab teams (Bernholdt et al., 2004), which targets high-performance computers and complex sophisticated models. All these frameworks are mostly concerned with the software side of the integration problem. The models are merely treated as software components that are to be made to work together and talk to each other.

From the software point of view there are several quite important problems to be solved when linking models together (Peckham, 2010):

- Models are written in different languages (conversion is time-consuming and error-prone);
- Code is not well-documented or easy to understand and reuse;
- Models have different geometry, dimensionality (1D, 2D or 3D);
- Models may use different types of grids (rectangles, triangles, polygons);
- Each model has its own time loop or "clock";
- Mismatched numerical schemes (explicit vs. implicit).

These problems are largely solved by the architectures mentioned above. However we should keep in mind that models are certainly more than just software, and as such linking them is associated with many more problems:

- Components are built by different teams, at different time, at different places. They are built for different goals and purposes, and it is not a given that they can be always reused ‘as is’;
- Teams use different languages, which makes it only harder to communicate the assumptions that the models build upon;
- In most cases approprieate metadata, metamodels and standards are missing or do not match;
- Modeling paradigms may be incompatible;
- Calibration is an important element of modelling and calibrating integrated models may be much harder than calibrating their builiding blocks;
- The time, space, and structure scales and resolutions may be very different and even incopatible;
- There is a propagation of error and uncertainties, which becomes only harder to track and account for as models become more complex. Increased complexity is almost inevitable when models are integrated.

Moreover, when integrating models, rather than software, we are more concened with integration of knowledge (that is embedded in each component) than only integration of code. This becomes especially evident if we recall that models may not necessarily be always quantitative. Some models are built in qualitative terms, and some are simply only conceptual. Is there any way that we can integrate such models and make them work together, sharing knowledge among various building blocks?

To start solving this problem we first need to make an itenerary of the models that are important for the COMPLEX project. We need to explore the project model-space and learn to characterize the various models in such a way that we can then pit them into some order. Just like books are put into different categories and catalogued in a library, similarly we need to find a way to describe the different models in such a way that we can then start making some connections between them and understand how they can be complementary, and how several of them can be used jointly to tell the story better.

Libraries of models are usually called *repositories*. We start with a brief overview of several existing repositories and look at their specifications. We then describe the COMPLEX project model space as it is today. We also realize that this model space is likely to change and we do need to have some flexibility in the way we describe models to accomodate the other ones that are likely to come up. However, the existing model space is important to understand to identify some of the basic dichotomies and categories that are already present and that we can start working with.

We then focus on one of the existing repositories, the one developed by the Community Surface Dynamics Modelling System (CSDMS). We consider it as one of the most appropriate candidates to build upon for our purposes. We intend to avoid the 'wheel reinvention' syndrome and whenever possible will use the existing tools instead of developing similar ones just to own them. Some preliminary agreements with CSDMS have been reached and they would be supportive of joining forces to improve the instruments that they have already developed if those do not yet match our needs. This is especially promising since all their software is open source and available for our scrutiny and extension.

We finally make some preliminary recommendations on how we think our work on standards and model specifications should develop. This is certainly an iterative process that should involve all modellers in the COMPLEX project, and, perhaps, beyond. We are especially interested in reaching out to modellers in the sister project ADVANCE.

2. Review of model repositories

Model repository is a place where models are put to make them available for sharing, collaboration, safe storage, etc. Heery et al. (2005) describes repositories as a 'collections of digital objects' that have the following characteristics: (1) content can be stored in a repository either by the content creator, owner or third party; (2) the repository architecture manages content as well as metadata; (3) the repository offers a minimum set of basic services e.g. put, get, search, access control; (4) the repository must be sustainable and trusted, well-supported and well-managed. Due to these characteristics repositories are different from other collections of digital objects such as directories, catalogues, etc.

Model repositories help to expedite knowledge exchange by enabling widespread use of models for research. Since models are stored with description, repositories assist in establishing a common standard for model description. Model repositories promote collaboration not only among model developers and model users, but also with journal and other scientific publishing venues (EMBL-EBI, 2013). It also encourages the availability of models with different versions as free software product.

Maintaining a model repository requires tasks like: managing hardware and software infrastructure, implementing model documenting standard (meta-model standard), storing models with associated resources, and managing intellectual property/licenses of model owners. Beyond ensuring the availability of models with the respective documentation, model repositories don't give warranty about the performance of the hosted models.

To increase usability and enhance collaboration a repository should: (1) consist of a document on 'how to use' the hosted models, (2) maintain data to initialize and run models (if needed), (3) support different model development methodologies, (4) allow different model development tools and techniques, and (5) allow models from several disciplines. Some of model repositories provide access to supercomputing, high-performance functionality, which makes them the preferred destination for models that require massive computations or parallel processing for execution.

Ideally, scientific model repositories expect that models are shard as open-source, using one of the respective licenses. The advantage of open source is that models that are made available as open-source "provide complete information transfer" (CSDMS, 2013). In addition, it is obvious that science advances through collaboration. In the modelling context, collaboration includes collective effort of individuals to design, code, debug, test, and document models. Generally when repositories are open to the public, models can be used in any manner the user would like to use them. The user can use them 'as is', or modify them, embed them inside a bigger system, distribute them commercially or for free. When modes are available as open-source other users can also help answer questions about them or can fix bugs in the model.

The other function of model repositories is to support model versioning. When a user of an open source model incorporates missing features, the latest version of the model is expected to be made available in the repository for the next users. To facilitate collaboration and to control change models need to be put under version control (Koegel *et al.*, 2010), and this can be achieved using respective software tools.

Depending on the objective of the owner of the repository different repositories can have different criteria to access and store models on them. The criteria could be: domain specific, methodology specific, development tool specific, peer-reviewed, etc. or any combination of them. On the other hand some of the repositories are free of such criteria.

The other difference between model repositories is the way they store models. Some of the repositories store models in a database, and others store models as files. The level of detail of model documentation is also very much differently among model repositories.

Let us next consider some examples that demonstrate how the above-mentioned requirements are imposed by different model repositories. Since we are exploring repositories in the context of COMPLEX model space, we would like to focus on repositories that are open to the public worldwide.

BioModels Database (http://www.ebi.ac.uk/biomodels-main/), at the European Bioinformatics Institute, is a model repository of computational models of biological processes. Currently it hosts more than 462 models. It hosts models described in peer-reviewed scientific literature and it allows scientific community to store, search and retrieve models of their interest. Models, their semantic annotation, and related information is stored in a set of MySQL tables which enables users to search not only for particular models based on their internal components elements, but also based on the extensive additional annotation. The requirement set by BioModels repository is models should be quantitative, domain specific, and peer-reviewed. Being repository of peer-reviewed models the abstract of the paper is also made available to users. The user can work locally by downloading the available resources, or can use available web services to access resources programmatically.

The **OpenABM** Consortium (http://www.openabm.org/), which is a node in the Computational Modelling for Socio-Ecological Science (CoMSES) Network, has a model library to preserve and maintain digital artifacts and source code of agent based models. Although the network mainly works to support and expand the development and use of computational modelling in the social and life sciences, the repository is open to models from any discipline. OpenABM aims to document models in a way a colleague is able to use it and derive similar results. The repository focuses on agent-based models, however it is open to a broader set of computational models such as geosimulation, cellular automata, dynamic networks, etc. When model owners upload their models they are asked to provide tags so that the search engine of the repository will be effective in performing relevant searches. Currently the OpenABM repository hosts 191 models. Even though it is not strictly restrictive the requirement here looks like methodology specific.

Physiome Model Repository (http://models.physiomeproject.org/) works with goal to provide a resource for the community to store, retrieve, search, reference, and reuse of CellML (XML based markup language) models. The repository is open to the public and currently contains more than 200 CellML files, together with associated metadata, citations, and figures. The major requirement to upload a model to Physiome repository is it should be built using CellML, which makes it both domain and tool specific repository.

NetLogo Model Repository (http://ccl.northwestern.edu/netlogo/models/index.cgi) is a repository and a multi-agent programmable modelling environment that hosts agent based models built using NetLogo. Models can come from any discipline but it should be agent based and programmed using NetLogo, which means the repository imposes both methodology and development tool as a requirement to store a model in it. When the user download the NetLogo programming tool all of the models (except models contributed by users) in the models library will be automatically included. Models contributed by users are available through the web. The different feature that we observe with NetLogo is that if a user doesn't want to upload her/his model to NetLogo repository, but still wants to make her/his model available through external website s/he can register the URL of the external website in NetLogo models list.

The Community Surface Dynamics Modelling System (**CSDMS**) (http://csdms.colorado.edu) model repository currently hosts 168 Earth's surface related models grouped under terrestrial, coastal, hydrological, marine, climate, and carbonate categories. To upload a model into CSDMS repository the model can be developed by any tool, any methodology, and actually from any discipline. We will describe this product in more detail in section 4.

Apromore (http://apromore.org/) - Advanced Process Model Repository (http://apromore.org/) is the result of collaboration among various universities for facilitating standardisation and reuse of business process best practices. The model repository is an open-source repository dedicated to store and to make available only business process models. Apromore repository allows hosted models to be developed by any tools and languages.

GitHub (https://github.com/) repository is one of the most popular repository by the scientific community and it works with motto 'build software together'. The repository is open not only for models but also for any software artefacts, about anything, and the product to be hosted could be either open-source or proprietary. Currently GitHub reports that it has more than 3 million registered users. In GitHub each uploaded software product acts as if it has its own repository with list of all its versions. And one of the special features of GitHub repository is its distributed version control. In distributed version control (Fig. 1) users will have full copy of the repository, and the version control system enables each of them to work in collaboration. However if any server dies, any of the client repositories can be copied back up to the server to restore it. One of the main shortcomings of GitHub repository is it didn't strictly enforce users to document metadata information which will have significant effect in searching and in getting description about the software product.

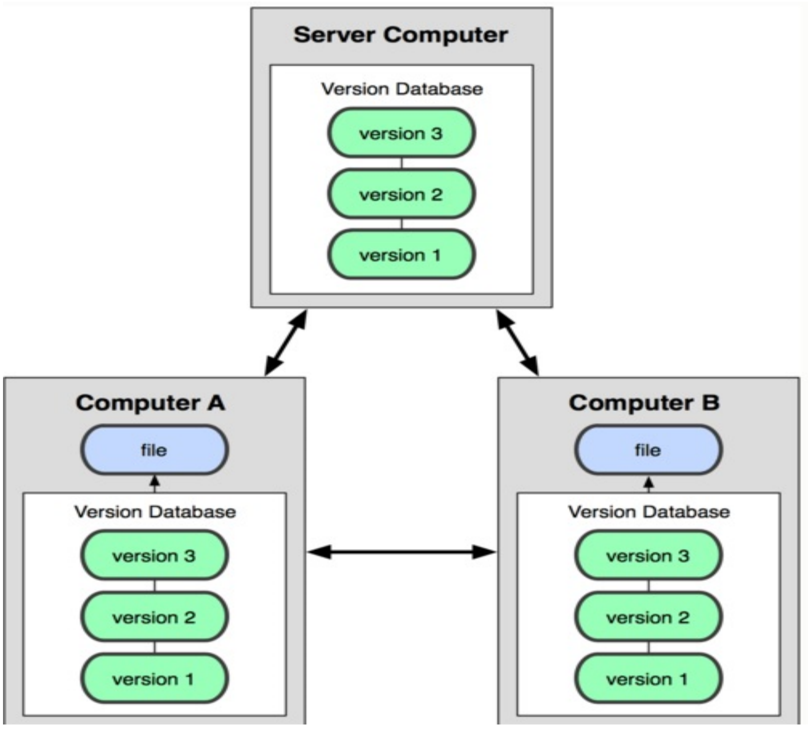

Figure 1. Distributed version control diagram (source http://git-scm.com/book/en/Getting-Started-About-Version-Control)

Model_repository SourceForge, which is a major open software repository, has a project (http://sourceforge.net/apps/mediawiki/peekabot/index.php?title=Model_repository) that is developing a model repository dedicated to robotics equipment models. For a model to be uploaded in Robotics repository it must be released under a free license that allows derivative works, redistribution and commercial use. The main criteria set by the repository is that the model should be both domain and methodology specific.

From the requirements set by model repositories we can observe that model repositories favour collaboration among scientists and contribute to the dissemination and continuity of knowledge. In addition, the availability of models in a repository can have its contribution to integration of models, since:

- model source code is available for integration framework developers,
- meta-model information can be reconstructed from the model source code and the model documentation of the repository,
- model users, other than model owners can easily contribute to testing the integration framework, which will increase scientific validity of the results of integration.

The COMPLEX consortium is established based on the concept of integrative research in inter-disciplinary areas. Having a model repository, which is accessible to all members of the consortium and beyond will enhance the collaboration level by one step. Based on the principles that drive our consortium we should expect that;

- the repository should be open to the public;
- should be inclusive for any discipline;
- should support a variety of modelling methodologies, tools, and techniques;
- should store models with documentation (meta-models); and
- should be well managed.

In this report we outline the framework for such a repository into which we can start uploading our models to make them available for future integration purposes.

3. COMPLEX model space

The COMPLEX model space (Fig. 2) consists of a number of models, which support research on climate change mitigation actions. As we can see in Table 1 it consists of models which are very diversified in different aspects: (1) model domains: climate, hydrology, land use, policy, and economy; (2) spatial characteristics: vary from global level to regional level and even having no spatial dimension; (3) temporal characteristics: range from yearly to hourly levels; (4) model type: most of them are quantitative and few are qualitative; (5) license type: some are open to the public and some are proprietary; (6) methodology used: agent based, system dynamics, cellular automata, etc.; (7) programming language used: Vensim, Netlogo, GAMS, Fortran, Matlab, C++, Fortran, etc.

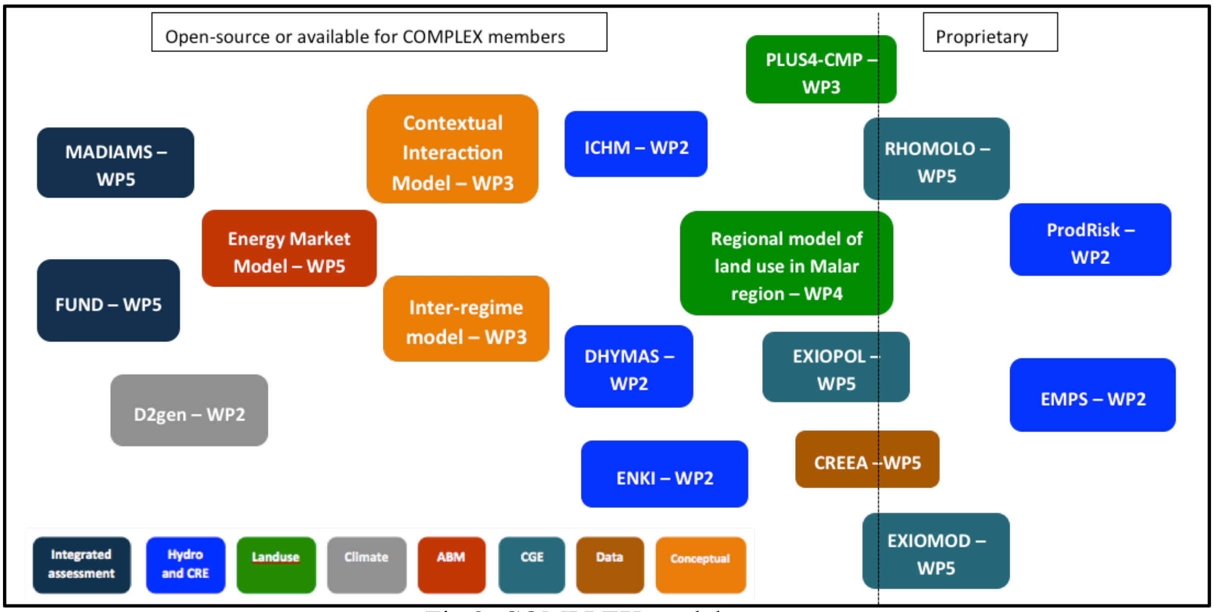

Fig.2. COMPLEX model space

Here we review some of the main models from our model space mainly with the perspective of model integration. The purpose of the review is to have a clear idea about the models and to identify which models (or their components) are available for integration. Based on this review we will; (1) identify possible coupling of models (components) among the COMPLEX models, and (2) we will develop use cases within each specific coupling. Having all these inputs will enable us to develop the prototype of our integration framework.

Keeping this in mind, we have found that most of the models are already built and some are in progress. Only a few models have ample description and in most of the cases we have been gathering pieces of information from websites, published papers, presentation slides, etc. which are not specifically targeted for model description.

Table 1. The COMPLEX Model Space

	Purpose, goal	Developer	License type	Software implementation / language used	Model type	Temporal charactristics		Spatial characteristics		Structural characteristics		
Model name		Contact person, e-mail			(Data, ABM, SD, stochastic, etc.)	Static/ dynamic	Time step (1 hour, 1 year)	Dimensions	Extent	Variables (units)	Inputs (units)	Outputs (units)
MADIAMS	Assessment of climate mitigation options	Klaus Haslemann (MPI-M), Dmitry Kovalevsky (NIERSC)	Open source (Vensim - proprietary)	Vensim	SD	D	1 year	0 - single region	single-region version - world economy (aggregated); multi-region versions (undedr development) - few macroregions (up to 10)	30 ODE	initial conditions (economy: GDP ($/ year), physical capital ($), stock of energy resources (mtoe); climate: CO2 concentration (ppm), CO2 emissions (GtCO2/year) etc.); parameters of actor control strategies	economy: GDP ($/ year), physical capital ($), stock of energy resources (mtoe); climate: CO2 concentration (ppm), CO2 emissions (GtCO2/year) etc.
energy market ABM	study non-marginal shifts in energy markets (including energy consumption and energy production) emerging from the bottom up (i.e. behavioral changes among households, diffusion of low-carbon energy sources)	Tatiana Filatova (t.fialtova@utwente.nl) with the support of other WP5 members	Open source	c sharp	ABM	dynamic, out-of equilibrium	multiple: (1) households agents and firms may be designed operating at a monthly time step, (2) while governments defining various policy instruments may operate at e.g. 1:5 years scale; (3) macro outcomes which are to be 'send' to CGE mdoels (EXIOMOD and Rhomolo) will most probably be at an annual basis	households and firms will be modelled at the NUTS2 or municipality scale. Specific location of agents within those regions will not be modelled.	either all EU or selected countries that correspond with other case-studies in the project	as the model is to be desined, we cannot say it now, but most likely many	survey data (e.g. on preferences of households regarding energy consumption), energy prices of the past periods, possible data on low-carbon energy diffusion, data on policy instruments for low-carbon energies	emergent: energy prices, possible socail norms regarding energy consumption
Contextual Interaction Model	Conceptualise the influence of the external context on local implementation processes through a set of predetermined actor characteristics	Hans Bressers j.t.a.bressers@utwente.nl	Open Source		Conceptual	D						
Inter-regime Model	Understand the joint influence of governance regimes on the local management of natural resources	c.deboer@utwente.nl	Open Source		Conceptual	S						
PLUS4-CMP	for participatory development of future land use scenarios for testing the impact/ implications in the territory of climate mitigation technologies (land use change)	richardjhewitt@gmail.com	some components are open source, but main cellular automata engine is proprietary	metronamica	geographical artificial intelligence,specifically cellular automata	Dynamic (stepwise, iterative)	1 year	2d cartesian plane (x,y)	various case study regions at NUTS 2	initial land use	institutional suitability (zoning), biophysical suitability, accessibility, neighbourhood rules, stochasitc factors, demand (land claims), this latter generated exogneously, perhaps through links to other project models (land prices, hydrology, energy demand etc)	final land use (future scenario maps)
d2gen	"Downscaling Disaggregation weather Generator". Disaggregation of GCM outputs (rainfall, temperature, wind, insolation) at finer time and space scales.	benoit.hingray@ujf-grenoble.fr	Available for COMPLEX partners	Matlab	Statistical downscaling model based on resampling. Combination of i) Generalized linear model and ii) K-nn resampling.	Preserves dynamics at infra daily scale as it is present in the data	Infra-daily. According to the regional dataset used.	2D space + time	Mesoscale domain (300x300 km2)		1- GCM outputs (pressure fields); 2- Regional data set of rainfall, temperature, wind, insolation at the space and time resolution the disaggregation is made	Rainfall, temperature, wind, insolation
Regional model of optimal land use applied to the Mälar region	Policy design for optimal land use achieving future targets on carbon	ing-marie.gren@slu.se	Could be open	GAMS	Stochastic (chance constrained programming)	dynamic	1 year	four spatial layers	Mälar region, approx. 30000 km2		land prices, ecosystem service value, carbon and nutrient sequestration, water quality, GHG emission, prices of output and inputs for land use, uncertainty quantification, risk aversion	Paths of optimal land uses in space and time, under different scenarios, which is used for identifyint optimal policies
ICHM	Hydrological rainfall runoff model at the catchment scale	Marco Borga <marco.borga@unipd.it>	Open Source		Deterministic	Dynamic	Infra-daily. According to the regional dataset used.	Catchment	1D + Time	Up to mesoscale	Precipitation and temperature. Catchment data.	Runoff (mm/h)

Model name	Purpose, goal	Developer Contact person, e-mail	License type	Software implementation / language used	Model type (Data, ABM, SD, stochastic, etc.)	Temporal charactristics Static/ dynamic	Temporal charactristics Time step (1 hour, 1 year)	Spatial characteristics Dimensions	Spatial characteristics Extent	Structural characteristics Variables (units)	Structural characteristics Inputs (units)	Structural characteristics Outputs (units)
DHYMAS	Hydrological rainfall runoff model at the catchment scale	Marco Borga <marco.borga@unipd.it>	Open Source	Fortran	Deterministic	Dynamic	Infra-daily. According to the regional dataset used.	Catchment and/or Drainage district	2D space + time	Up to mesoscale	Regional data set of rainfall, temperature, wind, insolation at the space and time resolution the disaggregation is made. District data. Irrigation.	Runoff (mm/h)
EXIOMOD	Computational General Equilibrium Model at the World level (95% world GDP)	Saeed Moghayer <s.m.moghayer@tno.nl> and Olga Ivanova <olga.ivanova@tno.nl>	proprietary	GAMS	Deterministic	Dynamic (inter-temporal)	Time step: 1 year	Regional level	Spatial: EU27+candidate member states+ world'd large emitters + rest-of-the-world country Sectoral: 129 Sectors (NACE classification)	Many! Can't be listed here but technical description is available	- EXIOBASE data set based on Environmentally Extended Input Output Tables (129 sectors * 43 countries + GHG and Non-GHG emissions + Land Use + Waste + Energy use) + other Economic data. Detailed Data set description is available. - Climate Senarios	Many consists of: (1) Social effects (2) Economic effects (3) Environmental effects. Detailed description is available.
RHOMOLO	Regional Computational General Equilibrium Model at EU NUTS2 Region	Olga Ivanova <olga.ivanova@tno.nl>	proprietary (TNO and IPTS)	GAMS	Deterministic	Dynamic (inter-temporal)	Time step: 1 year	Regional-level	All EU NUTS2 Regions	Many! Can't be listed here but technical description is available	Regional data (EUROSTAT and other sources and sectoral) + National accounts + Regional environmental data	Many consists of: (1) Social effects (2) Economic effects (3) Environmental effects. Detailed description is available.
EXIOPOL	DATA - Environmentally extended national accounting system	Olga Ivanova <olga.ivanova@tno.nl> and Saeed Moghayer <s.m.moghayer@tno.nl>	proprietary (EU Commission DG Research,TNO and Partners in EXIOPOL FP6 project)	NA	Spatial: 43 Countries (Incl. EU 27, Candidate member states, and large emitters) covering 95% World's GDP Sectoral: 129 Sectors (NACE classification)		NA	Country-level	43 Countries	Many! Can't be listed here but technical description is available	National Supply-Use tables	EE-Extended IO tables
CREEA	DATA - Environmentally extended national accounting system (Extension of Exiopol)	Tatiana Bulavskaya <tatiana.bulavskaya@tno.nl>	proprietary (EU Commission DG Research, TNO and Partners in CREAA FP7 project)	NA	Spatial: 43 Countries (Incl. EU 27, Candidate member states, and large emitters) covering 95% World's GDP Sectoral: 200 Sectors (NACE classification)		NA	Country-level	43 Countries	Many! Can't be listed here but technical description is available	National Supply-Use tables	EE-Extended IO tables
ENKI	A modular framework for building hydrological or other environmental models (temporal GIS)	Kolbjørn Engeland kolbjorn.engeland@sintef.no	Open Source / LGPL	Coded in Visual C++	Stocastic and deterministic models might be combined. Any model that runs on a spatial grid or network might be implemented.	Dynamic	Temporal resoultion adapted to process models and data used within the framework. Typically 1 hour, 1 day or 1 month	2D, can be defined on a regular grid or a set of nodes/points	Spacial resolution adapted to the feasable process description and case to be studied. Typicalli 200x200 km	Many! Can't be listed here but technical description is available. Extensive use of land surface maps (vegetation and soil characteristics)	Depends on processes included. Typically land surface weather (temperature, precipitation, humidity, radiation, wind)	Depends on processes included. Typically streamflow, soil moisture, snow water equivalent, ground water storage
ProdRisk	Hydro power scheduling	SINTEF Energy Research, Geir Warland, geir.warland@sintef.no	Proprietary	Coded in Fortran	Stochastic model		Basic time step one week, duration curve within week. Newly developed functionality allows finer time resolution within week.	Regional level	Regional level	MW, MWh and Mm3	Regional data set of inflow, temperature, wind. Market price forecasts.	Time series: hydro power schedules, reservoir levels, spillage. MWh and/or Mm3.
EMPS	Power market simulator, hydropower scheduling.	SINTEF Energy Research, Geir Warland, geir.warland@sintef.no	Proprietary	Coded in Fortran	Combination of stochastic and simulation model		Basic time step one week, duration curve within week. Newly developed functionality allows finer time resolution within week.	Multiple connected regions or areas (representing multiple market areas).	Multiple connected regions or areas (representing multiple market areas). Can represent multiple interconnected power markeds, e.g. Nordic area or Europa.	MW, MWh and Mm3	Regional data set of inflow, temperature, wind. Market data. Thermal production, demand: firm and price dependent.	Time series: hydro power schedules, reservoir levels, spillage. MWh and/or Mm3. Price forecasts and economic results.
FUND	Integrated assessment model for climate change		Open	C#.Net	Deterministic	Dynamic	Time step: 1 year	Regional level	16 defined regions exist			

EXIOMOD

EXIOMOD is Computable General Equilibrium (CGE) based model that tries to combine economic, environmental and social domains. It is mainly designed to perform long-term forecasting and impact assessment of policies at country and European level. It also tries to associate all production and consumption activities with the respective emission. The model uses actual economic data to estimate how an economy might react to changes in policy, technology or other external factors. EXIOMOD is dynamic and recursive over time, which is composed of a sequence of static equilibria that are connected to each other. EXIOMOD enables to study dynamics of capital accumulation and technology progress, stock and flow relationships and adaptive expectations.

EXIOMOD uses the notion of the aggregate economic agent to represent the behaviour of the whole population group or of the whole industrial sector as the behaviour of one single aggregate agent. EXIOMOD model economic agents represent micro-economic behaviour of: households, production sectors, investment agent, federal government and external trade sector. The agents are built by assuming that their behaviour is driven by certain optimization criteria such as maximization of utility of the households or cost-minimizing behaviour by producers and average-cost pricing.

Geographically the model incorporates the representation of 43 countries of the world, which includes an individual representation of all EU27 countries and candidate member states. It also includes the largest emitters such as US, Japan, Russia, Brazil, India and China. Countries, which are not represented separately in EXIOMOD, are grouped together into the Rest of the world "country".

EXIOMOD is proprietary model built using GAMS programming language. It works spatially at country level and temporally on yearly time scale. EXIOMOD uses a database (from EXIOPOL project) that contains data on 129 sectors and commodities.

RHOMOLO (Regional Holistic Model)

RHOMOLO is built for the ex-ante EU Cohesion Policy (ECP) impact assessment and also for ex-post impact assessment, other policy simulations and for comparison between the policy scenarios (Ferrara *et al.*, 2010). Hence RHOMOLO aims to be 'holistic' it tries to integrate the economic, environmental and social dimensions in one framework.

RHOMOLO belongs to the family of CGE models; and constructed using the concept of Dynamic Spatial Computable General Equilibrium (DSCGE)(Gardiner *et al.*, 2011). RHOMOLO is a multi-regional and multi- sectoral dynamic general equilibrium model that focus on EU NUTS2 (NUTS1 for Germany) region.

Each region in RHOMOLO is represented by economic agents such as: households, production sectors, regional and federal government. Being multi-regional and multi-sectoral models it tries to includes interregional trade and migration. In RHOMOLO interregional trade can takes place between the regions within a country or between the regions of two different countries; but interregional migration is assumed to take place only within the same country.

With regards of sectors, according to Eurostat classification RHOMOLO incorporates 23 types of sectors with the assumption each sector produces only one types of goods or services. Production and consumption by those sectors is associated with air pollution and generation of waste. For each types of greenhouse gas and non-greenhouse gas emissions RHOMOLO tries to quantify the damage in terms of monetary value.

Since RHOMOLO is a dynamic model it allows for the analysis of each period of the simulation time horizon. For each year of the time horizon, RHOMOLO calculates a set of various economic, social and environmental indicators. Time periods in RHOMOLO are linked by savings and investments; and the assumption is by the end of each time period, households, firms and government in the model save a certain amount of money.

RHOMOLO consist of modules like: Households' behaviour, Production sectors' behaviour, Multi-level government system, Interregional and international trade, Net regional migration, Investment and savings, Emissions and waste, Labour market. RHOMOLO is proprietary model built using GAMS modelling system, and operates on yearly temporal scale.

EXIOPOL (A New Environmental Accounting Framework Using Externality Data and Input-Output Tools for Policy Analysis)

EXIOPOL is an integrated Environmentally Extended Input–Output (EE I-O) database built to support cost-effectiveness and cost-benefit analysis of technologies, policies, and standard setting, at the micro (e.g. company), meso (e.g. sector) and macro (e.g. country or multi-country) (Tukker et al., 2009). The three principal objectives of the EXIOPOL was: (1) to synthesize and develop comprehensive estimates of the external costs of a broad set of economic activities for Europe; (2) to set up a detailed EE I-O framework with links to other socio-economic models; (3) to apply the results of the external cost estimates and EE I-O analysis for the analysis of policy questions.

EXIOPOL cover about 130 sectors and products. The database represents the EU27 and 16 non-EU countries such as US, Japan, China, Canada, South Korea, Brazil, India, Mexico, Russia, Australia, Switzerland, Turkey, Taiwan, Norway, Indonesia and South Africa. These countries were selected because they account for over 90% of global GDP and 80–90% of the trade volume by value with Europe.

EXIOPOL aims to evaluate, analyse, and assess damages from the emissions of pollutants into air and water. It tries to measure the monetary value of environmental burdens. The costs can be presented by type of emissions, industry sector, and country. EXIOPOL is proprietary product that functions at country level.

CREEA (Compiling and Refining Environmental and Economic Accounts)

The main goal of CREEA is to refine and elaborate economic and environmental accounting principles, to test them in practical data gathering, to troubleshoot and refine approaches, and show added value of having such harmonized data available via case studies. The main areas of study are waste and resources, water, forest and climate change (or Kyoto accounting). This includes work on developing harmonized data sets for integrated economic and environmental accounting.

Most data gathered in CREEA will be consolidated in the form of Environmentally Extended Supply and Use tables (EE SUT) and update and expand the EXIOPOL database. In this way, CREEA will produce a global Multi-Regional EE SUT with a unique detail of 130 sectors and products, 30 emissions, 80 resources, and 43 countries plus a rest of world. The ultimate contribution of CREEA is macro-database with supply and use tables for global monetary, physical and energy. And this will enable to build time series data that can calibrate models.

FUND (Climate Framework for Uncertainty, Negotiation and Distribution)

FUND, which is also called "Integrated assessment model of climate change" (Anthoff *et al.*, 2010), is a model that tries to show the impact of climate change on the economy. Originally FUND was developed to study the role of international capital transfers in climate policy, and it is now often used to perform cost-benefit and cost-effectiveness analyses of greenhouse gas emission reduction policies, to study equity of climate change and climate policy, and to support game-theoretic investigations into international environmental agreements.

Being an integrated assessment model it links scenarios and simple models of population, technology, economics, emissions, atmospheric chemistry, climate, sea level, and impacts. FUND tries to quantify impacts of climate change by setting monetary equivalent for: the premature death of individuals due to temperature stress or vector-borne diseases; migration of people due to sea level rise; the effect of climate change in agriculture, forestry, hurricanes, energy, water, ecosystems, etc. In FUND, some of the impacts of climate change are assumed to depend on the impact of the previous year. The initial year for the model is set to 1950, and based on this the model can run in time-steps of one year from 1950 to 3000.

FUND represents the world as 16 major regions, that is the United States of America, Canada, Western Europe, Japan and South Korea, Australia and New Zealand, Central and Eastern Europe, the former Soviet Union, the Middle East, Central America, South America, South Asia, Southeast Asia, China, North Africa, Sub-Saharan Africa, and Small Island States. It is an open source product in which the source code, data, and a technical description of the model can be found at http://www.fund-model.org. Version 3.7 is developed using C#.Net and data is available in Ms-Excel format.

MADIAMS (A Multi-Actor Dynamic Integrated Assessment Model System)

MADIAMS is an integrated assessment model built by coupling the climate sub-system (NICCS) with economic model (MADEM) with the objective of studying the interactions between climate and the socio-economic system. It is a system dynamics model that simulates the evolution of the economy by using the interactions of a few key aggregated actors (Hasselmann *et al.*, 2012). The model enables to investigate the net impact of the different responses of firms, shareholders, workers, consumers and banks to government climate mitigation policies such as a carbon tax and subsidies for investments in renewables, etc. And MADIAMS can present the impact of such measures in terms of key economic variables such as GDP, wages, unemployment, consumption, savings, etc.

The model treats the economy as a nonlinear system described by a set of system-dynamic equations. Structurally MADIAMS is an integral system that has three levels of hierarchy (Fig. 2) : (1) the lowest model level M1 is an economic system governed by the strategies of only three actors: firms, households and a bank. (2) The next level M2 includes governments as a fourth actor, and (3) the third level M3, tries to have complete coupled view of climate-socio-economic system.

MADIAMS is an open source product developed by using proprietary tool known as Vensim DSS Programming language. The model operates on yearly temporal scale; and spatially at single-region (global-scale) level and multi-region level (with the world disaggregated in few macro-regions).

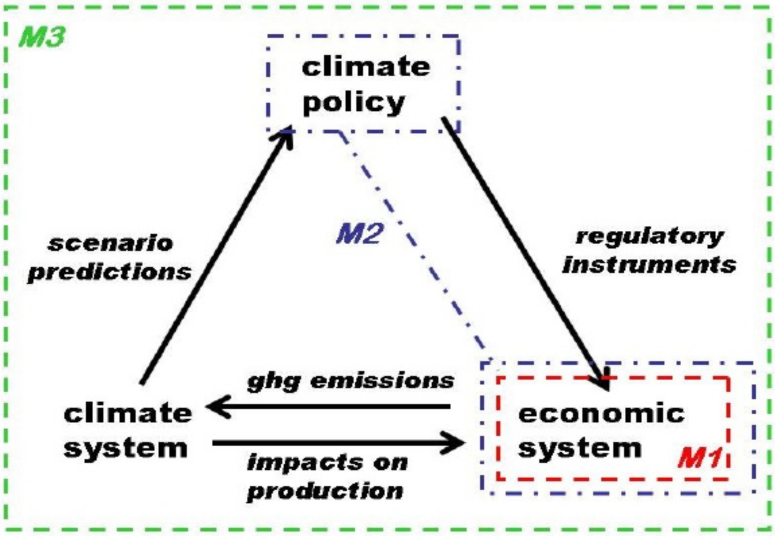

Figure 3. the three levels of MADIAMS: M1, M2, M3 and their interactions. (source: (Hasselmann *et al.*, 2012))

DD-WGEN (Downscaling-Disaggregation Weather Generator)

DD-WGEN is combined downscaling and disaggregation weather generator developed for multisite generation of hourly precipitation and temperature time series over complex terrain from large scale atmospheric information (Mezghani et al., 2009). DD-WGEN combines two kinds of statistical downscaling models (SDSMs) to produce weather scenarios for the present climate. The first is Generalized Linear Models (GLMs) for the generation of daily regional weather variables, and the second is a K-nearest neighbours resampling approach (K-nn) for their disaggregation to multisite hourly data. Although many hydrological applications require smaller time step, GLMs are commonly developed to generate weather variables at a daily time step. This arouses the need for temporal disaggregation methods.

DD-WGEN then uses the spatial (2D) and temporal pattern of the same day for the disaggregation of both regional variables. DD-WGEN is developed using MATLAB and available for COMPLEX partners. DD-WGEN was applied to the study area (Upper Rhone River basin) and had successfully reproduced standard statistics for temperature as well as total and liquid precipitation at the temporal and spatial resolutions required for hydrological modelling of the system (3 h, 100 km^2) and at lower resolutions down to those relevant at the river basin scale (3 days, ~5500 km^2). Although DD-WGEN works at different spatial and temporal scale it is observed that:

(1) the accuracy of statistical relationships for the prediction of precipitation from atmospheric predictors was found to be better at the regional than at sub-regional scales.
(2) the best performance is generally obtained for daily regional variables while performance tends to decrease for smaller spatial and time scales.

Hydrological Model - DHYMAS and ICHM

The hydrological model is aimed to evaluate a threshold-based flash flood warning method, by considering a wide range of climatic and physiographic conditions, and by focusing on ungauged basins (Norbiato *et al.*, 2008). The model is a semi-distributed conceptual rainfall–runoff model which is routinely used within the flood forecasting system in some river systems. It is built with the assumption that says, on a certain basin the depth of rain in a given duration is uniform in space and time, and this can cause minor flooding at the outlets of that specific basin. The model consists of snow routine, a soil moisture routine and a flow routing routine.

The model is an open-source model developed using Fortran programming language and it runs on an hourly time step. To run the model it requires 14 parameters as an input: three for the snow accumulation and melt module, eight for the probability distributed moisture module, and three for the runoff propagation module. Runoff in mm/h is the main output of the model.

The model was tested on two distinct European regions: northeastern Italy (with eight basins) and central France (with three basins). The result found shows that the model can be applied for flood warning systems, with focus on ungauged basins. In COMPLEX project the model will be used to generate the inflow time series to the hydropower reservoir.

ENKI

ENKI is an Open Source (http://www.opensource-enki.org/) modular platform for hydrological model implementation (Kolberg et al., 2012). Originally ENKI was developed for the purpose of improving hydrological forecasting for hydropower scheduling. ENKI consists of a suite of separately compiled subroutine modules that enables to build space-time models for hydrological or other environmental purposes. Being an open-source product and having modular design has enabled ENKI to be rapidly disseminated for operational hydropower forecasting or other water resource management purposes.

So ENKI consists of a set of user-defined subroutines, and operates on GIS data within a geographical region. Temporally it can work on an hour, a day, or a month scale; and spatially typical resolution scale is 200x200km. ENKI is developed using C++ and uses a plug-in structure to build a complete model from separately compiled subroutine implementations which are compiled as dynamic-link libraries (dll).

ProdRisk

Like any other business hydro power production have different kind of risk like: price risk, quantity risk - caused by inflow uncertainty, quantity risk - caused by demand uncertainty, etc. To manage such kinds of risk we need risk management tools, and ProdRisk is one of such tools developed for Scandinavian market.

ProdRisk is a model for optimisation and simulation of hydro-thermal systems. It uses stochastic dual dynamic programming to solve the optimisation problem. It is an integrated model used for long term planning and seasonal planning of hydropower plants (Mo *et al.*, 2001). ProdRisk also includes functionality for integrated risk management and dynamic hedging in the futures market.

ProdRisk can be used for medium and long-term hydro power scheduling (i.e. 2 to 5 year time horizon) of local or regional energy-systems. For analysis the user can define time-step on hourly, daily or weekly basis. Inflows to the reservoirs, temperature, wind, and market prices for electricity (i.e. if market prices are modelled externally) are main stochastic inputs for ProdRisk. Scenarios for reservoir operation, hydro production, marginal value of water in different reservoirs and a profit distribution are outputs. ProdRisk is proprietary software developed using Fortran programming language.

EMPS

Depending on the situation we can use different source of energy for the production of the required amount of energy by the power market. On the other hand renewable energy sources like wind power, solar power, etc., can only be produce whenever the energy source is available. Which means our system should have backup power for periods when such energy sources are not available. Due to this we need a mechanism to schedule energy production level to address the demand of the power market.

Based on this, EMPS is Multi-area Power-market Simulator designed for long and medium term scheduling of hydrothermal electricity systems (Warland *et al.*, 2011). The EMPS model has been used to calculate the cost minimizing operation of a hydro-thermal power system in the Nordic power market (Wolfgang *et al.*, 2009). The model may include more than 1000 hydro reservoirs, several hundred thermal plants, wind power and solar power described with hourly production data etc. The Nordic countries is widely using EMPS for price forecasting, generation scheduling, expansion planning and general system analyses.

EMPS uses stochastic dynamic programming (SDP) algorithm for the hydro scheduling part. The basic time resolution of the EMPS-model is 1 week with a load duration curve within each week, but the newly developed functionality allows finer time resolutions within a week also. The model takes inputs like regional data set of inflow, temperature, wind, market data, and thermal production; and produces time series hydropower schedules, reservoir levels, spillages, price forecasts and economic results as an output. EMPS is proprietary model developed using Fortran programming language.

Energy Market Model

It is an agent based energy market model with the goal to study non-marginal shifts in energy markets (including energy consumption and energy production) emerging from bottom up (i.e. behavioural changes among households, diffusion of low-carbon energy source) . The model is under development and it will be developed as an open source product.

Temporally the system will function at multiple states: (1) households agents and firms may be designed operating at a monthly time step; (2) while governments defining various policy instruments may operate at e.g 1:5 years scale; (3) macro outcomes which are to be ‘send’ to Climate Related Energy (CRE) models (like EXIOMOD and RHOMOLO) will most probably be at an annual basis. Survey data like preferences of households regarding energy consumption, energy price of the past periods, possible data on low-carbon energy diffusion, data on policy instruments for low-carbon energies will be used as an input by the model. Energy prices, possible social norms regarding energy consumption are the expected outputs of the model.

PLUS4-CMP (Participatory Land Use Simulator for Climate Mitigation Policies)

PLUS4-CMP is aimed to simulate the complex interaction as a result of policies intended to reduce carbon emission. These interactions mainly influence economy, land-use patterns, social cohesion and compliance. PLUS4-CMP tries to identify potential social and ecological conflicts due to renewable energy expansion. The two main modules of PLUS4-CMP model are: land demand for renewable energy uses, and land use allocation for renewable energy. The land demand module determines the amount of land to be used for renewable energy production. It tries to answer how much land is necessary for renewable energy implementation in each different scenario. On the other hand the land use module determines which scenarios are going to be simulated. It tries to handle the interaction of three key elements: where land will be allocated, how will this allocation occur, and how much land is required. Currently PLUS4-CMP is under development and it aims not only to develop tools and approaches that explore issues related to renewable energy implementation but also policy recommendations on the basis of analysis. The model will work at yearly time step up to year 2050.

4. The CSDMS repository

The Community Surface Dynamics Modelling System (CSDMS) is a diverse community of experts that deals with the Earth's surface modelling. CSDMS maintains a large and searchable repository of contributed models (Peckham *et al.*, 2013). Currently the CSDMS repository hosts 168 models categorized as Terrestrial, Coastal, Hydrological, Marine, Climate, and Carbonate.

As the word “Community” in its name indicates, CSDMS aims to improve collaboration in coding, debugging, testing, documenting, and using of models and modelling frameworks. CSDMS states that beyond collaboration “research should be public and community centered as well” (CSDMS, 2013) and this can be achieved by following the open-source standard. In addition to hosting models and modelling frameworks CSDMS has a high performance computing cluster (HPC), which is free of charge to the CSDMS community.

To submit a model to CSDMS repository the model owner is expected to do three things: choose license type from available types of licenses, fill questionnaire about the model, and compress the source code and submit it. Optionally, if the model owner wants to make her/his model to be part of CSDMS compliant plug-and-play model component s/he has to: implement a Basic Model Interface (BMI) to the model, and use CSDMS standard names that map input and output variable names. Currently the BMI can be implemented for only models developed using C, C++, Fortran (all years), Java and Python; other programming languages are not supported.

To host COMPLEX member models in the CSDMS model repository we were having discussions with the administrator of the repository and we have come up with the following issues:

- Even though CSDMS is mainly focused on Earth surface models, they are currently expanding to also represent the ‘anthroposhere’. Therefore CSDMS is quite interested in exploring how socio-economic models can also be part of their system and therefore COMPLEX member models are most welcome to be hosted in CSDMS model repository as long as model owners agree to follow open-source standards.

- The COMPLEX models are developed with very diverse tools (beyond C, C++, Fortran, Java and Python), which cannot implement CSDMS BMI interface. CSDMS has agreed that we can implement our own model componentization and integration methodology in the repository.

- CSDMS can grant access to the HPC cluster, but for security reasons access over internet is not allowed.

- Since COMPLEX involves a variety of models the meta-model information documented using CSDMS questionnaire may not be sufficient for our integration need. However they have agreed to consider our suggestion on meta-model information questioner if they found it sounding. On the other hand we can leave CSDMS questionnaire as it is and we can use our meta-model documenting template internally among COMPLEX modellers.

Generally, as mentioned above in the model repository section, having a central repository for models can be instrumental for the integration process and can also boost the overall collaboration among the consortium members. The information we gathered shows that the CSDMS repository can be sufficient for hosting COMPLEX member models, so we could better use our resources if we team up with CSDMS instead of re-inventing wheels and creating yet another specialized model repository. We would recommend that CSDMS repository be used, however we do run into some problems, primarily regarding the model licensing and access.

Since CSDMS is largely funded by federal money (NSF) it is specifically restricted for open-source models only. However the COMPLEX model space shows that some of the models are proprietary, and some of the potentially open source models are still not ready to be shared beyond COMPLEX. Due to these reasons in COMPLEX we will have to design two-layered framework, which will consist of a unified data-base for model documentation (meta-model) with further links, either to the CSDMS repository for open-source models or directly to individual web-sites that will host the proprietary models or 'models under development'. As a first step we will primarily focus on documenting our models for collaboration and will be less concerned about the actual running of these models in an automated fashion.

5. Standards and meta-models

As scientific databases continue to grow in volume, breadth and complexity better descriptions of data (i.e., metadata) are essential for understanding and using the increasingly complex and voluminous data and information (Michener, 2006). Similar to metadata, we imply that a meta-model is a simplified description of a model, a model of a model. It describes model assumptions, its constituents (variables, parameters, relationships), and its input-output (forcing, control functions, etc.). A meta-model tries to describe how a subject is represented using the model (Darnton, 2012) (Fig. 4). Like metadata, which are data about data, a meta-model is a model of a model and describes the 'who, what, when, where, and how' aspect of the model. So a meta-model describes the model to its users and helps to indicate possible collaborations or extensions and connections of the model with other models and/or data.

Meta-models can be developed for a wide range of purposes. They could be just giving a brief description of the model, or inform the user about how s/he can extend the model. Like with any model, depending on the purpose of the meta-model the level of detail of meta-models is also widely varied. Despite the importance and large potential of meta-modelling there is only limited progress in developing meta-model standards so far. Meta-model development has been highly dependent on the developer.

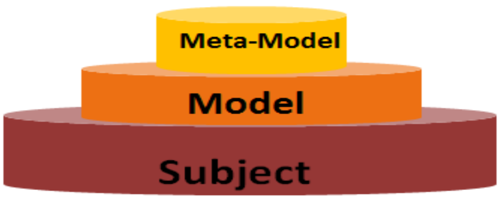

Fig 4 Meta-model layer

To build integration of models the meta-models of participating models play a crucial role since they document basic features and contextual information of the models. Creating meta-models is one of the basic milestones in developing the integration framework. It can also serve as reference for understanding the integration framework for future applications for new model integrators. Kuhn et al. (2003) point out that for integration of models that describe interdisciplinary concepts "a necessary prerequisite is the integration of their underlying metamodels" (Kuhn et al., 2003). In integrating models, if the context of the first model is not understood "the transferred part misbehaves" (Benz *et al.*, 2001) in the second model, and the result could be worse if it goes unnoticed. In integrating models building meta-models can be seen as the starting point of collaboration.

We have reviewed several existing meta-model standards that can be useful for the COMPLEX model space. Currently we can find a number of metadata standards, as listed in (Riley, 2010), e.g. DIF (Discovery Interchange Format) – metadata for description of scientific data sets; EML (Ecological Markup Language) - metadata for ecological information including raw data, published research papers, rights information, and research protocols; GEM (Gateway to Educational Materials) – metadata for the description of educational resources, etc. However there is no generally accepted meta-model standard so far. Meta-model standards are still mostly under development. One of the main reasons that we do not have universally accepted meta-model standards (while we have standards for metadata) is because there are many more modelling techniques and paradigms than what we find in data collection. For clarification let us consider some of implemented meta-model templates.

The **Reusable Asset Specification** of OMG (Object Management Group) (OMG, 2013) has guidelines and recommendations about the structure, content, and descriptions of reusable software asset documentation. The scope is defined for many software artefacts, which can be any work products from the software development lifecycle, such as requirements documents, models, source code files, deployment descriptors, test cases or scripts, and so on. So it includes features which are not limited to metadata of models. While useful, it is still focused largely on the software aspects of modelling, but not models themselves.

The **Ecological Modelling System**, ECOBAS (Benz et al., 2001) has specification for model documentation. It underlines that modelling and model documentation should go hand in hand, and recommends that no matter how large a model is, its documentation should give enough information to run it. But we also find that the ECOBAS model documentation strategy is mainly focused on source code documentation.

ISO/IEC 19763, **Information Technology Meta-model Framework for Interoperability** (MFI) has meta-model registering mechanisms that "allows the registration of various types of models, modelling constructs and concept scheme such as ontology" (Horiuch et al., 2013). The objective of the framework is to improve interoperability of meta-models defined by different standards groups in ISO or outside ISO. However, standardization of the contents of the meta-model is not covered, which makes it not applicable for our needs.

ODD (Overview, Design concepts, and Details) protocol (Grimm *et al.*, 2010) is a known and quite widely accepted meta-model standards for agent based models with the aim "to make model descriptions more understandable and complete". The ODD has seven elements (Table 2) named as: 1. Purpose; 2. Entities, state variables, and scales; 3. Process overview and scheduling; 4. Design concepts; 5. Initialization; 6. Input data; and 7. Sub-models. Since ODD is primarily developed for individual based and agent based models it has sub-elements (under Design concepts element) like emergence, adaptation, learning, prediction, sensing, interaction, etc. which could be irrelevant for agent-less models. In addition ODD lacks features like model output and some contextual information which are very useful in integration of models. On the other hand ODD + D (Müller *et al.*, 2012) is introduced to incorporate 'how human decision-making has been modelled' in the meta-model of agent based social science models.

	Elements of the updated ODD protocol
Overview	1. Purpose 2. Entities, state variables, and scales 3. Process overview and scheduling
Design concepts	4. Design concepts • Basic principles • Emergence • Adaptation • Objectives • Learning • Prediction • Sensing • Interaction • Stochasticity • Collectives • Observation
Details	5. Initialization 6. Input data 7. Submodels

Table 2 Elements of the ODD protocol

Besides the above mentioned standardization efforts, model repositories are using their own in-house developed meta-model templates for documenting models. As a start for our integration practice, by analysing our metadata requirements we have developed a draft meta-model documenting template, which is designed mainly for integration. The ODD and CSDMS meta-model documenting templates have served as a basis in developing the template.

Below in Table 3 we have sketched a preliminary draft of a meta-model template that we can use in the COMPLEX model space. However to become functional and accepted by the community of modellers in COMPLEX and beyond we need to engage in an iterative process of discussion and revision.

In our template a model may consist of a number of components (modules), and the template is designed to document models in terms of components (modules). Components (modules) that will be involved in the integration have to be documented using the template. The more components (modules) we can document the more we will learn from the process and the more material we will gather for future integration. During integration, the information gathered using the meta-model template will serve in establishing interoperability among different models, in semantic mediation between multidisciplinary models, and in converting the output of one model to input of another model.

Model Name:
Name of Component (module):
• Domain of component(module):
• Purpose of component (module):
• Component (module) key words:
• Model type ○ System Dynamics ○ Agent Based ○ Empirical ○ ANN ○ Bayesian ○ Conceptual ○ CGE ○ LCA ○ Data ○ Other
• Modelling techniques (mathematics) ○ Linear algebraic equations ○ Differential ○ Partial derivatives ○ Statistical ○ Logical ○ Other
• Model assumptions
• Spatial dimensions (more options possible): ○ No spatial dimension ○ 1D ○ 2D ○ 3D

• Spatial extent (more options possible): ○ Global ○ 1,000 km ○ 100 km ○ 10 km ○ 1 km ○ 100 m ○ 1-10 m ○ 0.00001 m – 1m ○ Point based (spatially homogeneous)
• Spatial resolution (more options possible): ○ 100km x 100km ○ 10km x 10km ○ 1km x 1km ○ 100m x 100m ○ 10m x 10m ○ 1m x 1m ○ 1cm x 1cm
• Temporal extent (more options possible): ○ No temporal dimension (static) ○ 100 years ○ 50 years ○ 10 years ○ 1 year
• Temporal resolution (more options possible): ○ No temporal dimension (static) ○ Year ○ Month ○ Week ○ Day ○ Hour ○ Minute ○ Second
• Description of input:
• Input entities ○ Name of entity: ○ Attributes (state variables) ▪ Name of attribute ▪ Data type ▪ Units used ▪ Valid range of input (add similar detail for all input entities)
• Input format (more options possible): ○ ASCII ○ Binary ○ Other

• Description of output (including intermediate variables):
• Output entities ○ Name of entity: ○ Attributes (state variables) ▪ Name of attribute ▪ Data type ▪ Units used ▪ Valid range of output (add similar detail for all output entities)
• Output format (more options possible): ○ ASCII ○ Binary ○ Other
• Supported platforms (more options possible): ○ Unix ○ Linux ○ Mac OS ○ Windows ○ Other platform
• Programming language used:
• License type (more options possible): ○ Proprietary ○ Open source
• Availability for integration:
• Can couple with (optional):
• Comment for integration (optional):

Table 3: Draft meta-model template for integration of models.

6. Conclusions

Being a place to store models and associated documents, repositories facilitate collaboration within the scientific community and beyond. Basic services of repositories like put, get, search, and access control assures the availability, sharing, and safety of the models paving the way for collaboration, joint discovery and further improvement of modelling tools. In addition to these basic services, other functionalities provided by repositories (e.g. meta-model documentation, version control, data for model initialization, and high performance computing) can further assist knowledge exchange and reproducibility of results. Integration of models requires collaboration in designing, coding, debugging, testing, and documenting the integration framework. Access to models through free and open-source repositories can be crucial for the success of the integration effort.

The COMPLEX model space consists of models from climate, hydrology, land use, policy, and economics domains. These models work at different temporal and spatial scales; are developed using different methodology, tools and techniques; are both qualitative and quantitative by type; and have both open and proprietary licensing. Understanding about the COMPLEX models will enable us to identify the possible models (components) for couplings and to develop integration use cases. This will further lead to the development of the prototype of our integration framework.

The CSDMS is one of the existing scientific communities that maintains a large and searchable repository of contributed models. Based on our preliminary assessment we found much potential in utilising the CSDMS repository for our purposes. We have a history of collaboration with CSDMS and find that they are also willing to adapt to our needs, which is very encouraging. Therefore we suggest that start using the CSDMS repository for hosting open-source based COMPLEX models as soon as possible, while we maintain the higher level meta-model data-base to provide information about and links to all the models in the COMPLEX model space.

Making models available on free and open repository will enhance collaboration and will add model transparency, access and reuse. However it is not sufficient to simply upload your model to the repository. The real bottleneck for model reuse and integration is the appropriate documentation of models in terms of standardised meta-model descriptions. So far we are seeing that the content of the meta-model templates is highly dependent on the purpose of model documenting. In our case the purpose is integration of models from a suit of very different disciplines, which makes their documentation only more difficult. We have developed a proposal for a meta-model template that we think can be instrumental for our goal of integration of models. This template will be used only if it is accepted by the majority of our collaborators. Therefore our next step is a comprehensive discussion of the meta-model template in search of a consensus that could lead to standard for interdisciplinary model declaration to be further used in the project. After that we should make sure that the models from the COMPLEX model space (and/or their components) are documented using the developed meta-model template and stored in the COMPLEX repository, which can be a sub-domain in CSDMS repository.

7. References

Ahuja, L.R., Ascough II, J.C., David, O., 2005. Developing natural resource models using the object modeling system: feasibility and challenges. Advances in Geosciences 4, 29e36. http://hal.archives-ouvertes.fr/docs/00/29/68/06/PDF/ adgeo-4-29-2005.pdf.

Argent, R., 2004. An overview of model integration for environmental applica- tionsdcomponents, frameworks and semantics. Environmental Modelling & Software 19 (3), 219e234.

Argent, R.M., Perraud, J.-M., Rahman, J.M., Grayson, R.B., Podger, G.M., 2009. A new approach to water quality modelling and environmental decision support systems. Environmental Modelling & Software 24 (7), 809e818.

Anthoff D, Tol RS (2010) FUND–Climate Framework for Uncertainty, Negotiation and Distribution.

Babel, 2004. Lawrence Livermore National Laboratory homepage. http://www.llnl. gov/CASC/components/babel.html.

Bicknell, B., Imhoff, J., Kittle, J., Donigian, A., Johanson, R., Barnwell, T., 1996. Hydrologic Simulation Program e FORTRAN User's Manual for Release 11. United States Environmental Protection Agency Environmental Research Laboratory, Athens, GA.

Barthel, R., Janisch, S., Schwarz, N., Trifkovic, A., Nickel, D., Schulz, C., Mauser, W., 2008. An integrated modelling framework for simulating regional-scale actor responses to global change in the water domain. Environmental Modelling & Software 23 (9), 1095e1121.

Bernholdt, D.E., Allan, B.A., Armstrong, R., Bertrand, F., Chiu, K., Dahlgren, T.L., Damevski, K., Elwasif, W.R., Epperly, T.G.W., Govindaraju, M., Katz, D.S., Kohl, J. A., Krishnan, M., Kumfert, G., Larson, J.W., Lefantzi, S., Lewis, M.J., Malony, A. D., McInnes, L.C., Nieplocha, J., Norris, B., Parker, S.G., Ray, J., Shende, S., Windus, T.L., Zhou, S., 2004. A component architecture for high performance scientific computing. International Journal of High Performance Computing Applications. https://e-reports-ext.llnl.gov/pdf/314847.pdf, ACTS Collection Special Issue, 75 pp.

Benz J, Hoch R, Legović T (2001) ECOBAS—modelling and documentation. *Ecological Modelling* **138**, 3-15.

CSDMS (2013) *The Community Surface Dynamics Modeling System (CSDMS).* http://csdms.colorado.edu/wiki/Main_Page

CMAQ, 2009. http://www.cmaq-model.org/ and http://www.chesapeakebay.net/committee_msc_projects.aspx?menuitem1⁄416525#peer

David, O., Markstrom, S.L., Rojas, K.W., Ahuja, L.R., Schneider, I.W., 2002. The Object Modeling System. In: Ahuja, L., Ma, L., Howell, T. (Eds.), Agricultural System Models in Field Research and Technology Transfer. CRC Press, pp. 317e330.

Darnton G (2012) *Meta meta languages and models.* http://metametametamodel.com/

EMBL-EBI (2013) *BioModels Database.* http://www.ebi.ac.uk/biomodels-main/

Ferrara A, Ivanova O, Kancs D (2010) Modelling the policy instruments of the EU cohesion policy. DG Regio working paper.

FRAMES, 2009b. http://www.epa.gov/athens/research/projects/3mra/index.html

Gaber, N., Laniak, G., Linker, L., 2008. Integrated Modeling for Integrated Environmental Decision Making EPA White Paper, 100/R-08/010.

Gardiner B, Brandsma A, Ivanova O, d'Artis K (2011) RHOMOLO: A Dynamic General Equilibrium Modelling Approach to the Evaluation of the EU's Regional Policies **11**, 672.

Grimm V, Berger U, DeAngelis DL*, et al.* (2010) The ODD protocol: a review and first update. *Ecological Modelling* **221**, 2760-2768.

Hasselmann K, Kovalevsky DV (2012) Simulating animal spirits in actor-based environmental models. *Environmental Modelling & Software*.

Heery R, Anderson S (2005) Digital repositories review.

Horiuch H, Gordon K (2013) *ISO/IEC 19763, Information Technology – Metamodel Framework for Interoperability (MFI).* http://metadata-standards.org/19763/

Koegel M, Helming J (2010) EMFStore: a model repository for EMF models, 307-308.

Kolberg S, Bruland O (2012) ENKI-An Open Source environmental modelling platfom **14**, 13630.

Kuhn H, Bayer F, Junginger S, Karagiannis D (2003) Enterprise model integration. In: *E-Commerce and Web Technologies, Proceedings* (eds. Bauknecht K, Tjoa AM, Quirchmayr G), pp. 379-392.

Kralisch, S., Krause, P., David, O., 2004. Using the Object Modeling System for hydrological model development and application. In: Proceedings of the iEMSs 2004 International Conference. University of Osnabrück, Germany. http://www. iemss.org/iemss2004/pdf/ integratedmodelling/kralusin.pdf, 6 pp.

Moore, R., Gijsbers, P., Fortune, D., Gergersen, J., Blind, M., 2005. OpenMI Document Series: Part A Scope for the OpenMI (Version 1.0). HarmonIT.

Mezghani A, Hingray B (2009) A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin. *Journal of Hydrology* **377**, 245-260.

Michener WK (2006) Meta-information concepts for ecological data management. *Ecological informatics* **1**, 3-7.

Mo B, Gjelsvik A, Grundt A (2001) Integrated risk management of hydro power scheduling and contract management. *Power Systems, IEEE Transactions on* **16**, 216-221.

Müller B, Angermueller F, Drees R*, et al.* (2012) Describing Human Decisions in Agent-Based Social-Ecological Models-ODD+ D an Extension of the ODD Protocol. *Available at SSRN 2044736*.

Norbiato D, Borga M, Degli Esposti S, Gaume E, Anquetin S (2008) Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *Journal of Hydrology* **362**, 274-290.

OMG (2013) *OMG Specifications*. http://www.omg.org/spec/

OpenMI, 2009. The Open-MI life project website. http://www.openmi-life.org/.

Peckham, S. 2010. CSDMS Handbook of Concepts and Protocols: A Guide for Code Contributors. http://csdms.colorado.edu/wiki/Help:Tools_CSDMS_Handbook

Peckham SD, Hutton EWH, Norris B (2013) A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Computers & Geosciences* **53**, 3-12.

Riley J (2010) *Glossary of Metadata Standards* J. Riley.

Tukker A, Poliakov E, Heijungs R*, et al.* (2009) Towards a global multi-regional environmentally extended input–output database. *Ecological Economics* **68**, 1928-1937.

Warland G, Mo B, Helseth A (2011) Handling balancing power in a power market with a large share of hydropower, 1-7.

Wolfgang O, Haugstad A, Mo B*, et al.* (2009) Hydro reservoir handling in Norway before and after deregulation. *Energy* **34**, 1642-1651.

Warner, J.C., Perlin, N., Skyllingstad, E.D., 2008. Using the Model Coupling Toolkit to couple earth system models. Environmental Modelling & Software 23 (10e11), 1240e1249.