Environmental Modelling & Software 52 (2014) 192–199

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Short communication

Documenting, storing, and executing models in Ecology: A conceptual framework and real implementation in a global change monitoring program

Francisco J. Bonet [a,b,*], Ramón Pérez-Pérez [a,b], Blas M. Benito [a,b], Fábio Suzart de Albuquerque [a,b], Regino Zamora [a,b]

[a] Laboratorio de Ecología (iEcolab), Instituto Interuniversitario Sistema Tierra, Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, Spain
[b] Departamento de Ecología, Facultad de Ciencias, Universidad de Granada, Avda. Fuentenueva s/n, Granada 18071, Spain

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form
23 October 2013
Accepted 26 October 2013
Available online

Keywords:
Algorithms
Metadata
Model
Documenting
Workflow

ABSTRACT

Many of the best practices concerning the development of ecological models or analytic techniques published in the scientific literature are not fully available to modelers but rather are stored in scientists' digital or biological memories. We propose that it is time to address the problem of storing, documenting, and executing ecological models and analytical procedures. In this paper, we propose a conceptual framework to design and implement a web application that will help to meet this challenge. This tool will foster cooperation among scientists, enhancing the creation of relevant knowledge that could be transferred to environmental managers. We have implemented this conceptual framework in a tool called ModeleR. This is being used to document, share, and execute more than 200 models and analytical processes associated with a global change monitoring program that is being undertaken in the Sierra Nevada Mountains (south Spain). ModeleR uses the concept of scientific workflow to connect and execute different types of models and analytical processes. Finally, we have envisioned the creation of a federation of model repositories where models documented within a local repository could be linked and even executed by other researchers.

Software availability

Name: ModeleR
Developer: Laboratorio de Ecología (iEcolab). Instituto Interuniversitario Sistema Tierra
Contact information: Avda. del Mediterráneo s/n, Granada 18006. Spain
Hardware required: General-purpose computer with Internet connection
Software required: Internet browser (later versions are recommended)
Program language: Ruby On Rails, C++
Availability: ModeleR is at http://modeler.obsnev.es. It is freely accessible after online registration. All the code needed to install ModeleR in a server can be downloaded following this GitHub link: https://github.com/rperezperez/modeler

1. Introduction: the challenge of storing, documenting, and executing models in Ecology

Data analysis, modeling, and simulation play a central role in Ecology. From statistical models to complex numerical models, the concept of the modeling and data processing has become inherent to the ecological research. This is due to at least three main factors.

First, there is a vast amount of primary information (data acquired by monitoring methods) available to modelers. This significant surge in data availability is a result of the enormous effort that scientists and public agencies have been exerting over the last 3–4 decades in order to gather and share information on the structure and functioning of the Earth's ecosystems. The creation of vast data infrastructures that allow environmental data sharing (Whitlock, 2011) (e.g. GBIF: Global Biodiversity International Facility, LTER: Long Term Ecological Research, DataONE: Data Observation Network for Earth, EarthCube) are greatly contributing to provide

* Corresponding author. Universidad de Granada, Instituto Interuniversitario Sistema Tierra, Laboratorio de Ecología (iEcolab), Avda. del Mediterráneo s/n, 18006 Granada, Spain. Tel.: +34 958249748; fax: +34 958137246.
E-mail address: fjbonet@ugr.es (F.J. Bonet).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.10.027

the primary data needed for Ecology to be fostered as a data intensive discipline (Jones et al., 2006; Kelling et al., 2009; Michener and Jones, 2012).

Second, the volume and availability of analytic and modeling methods have increased exponentially in recent decades (Crowley, 1992; Green et al., 2005; Metzger et al., 2011). Thanks to this trend, modelers can choose among dozens of different methodologies to analyze the structure or functioning of a given natural system, and to design complex models to simulate it under different scenarios.

Third, available software and hardware allow scientists to model complex systems and cope with data intensive computing procedures (Fegraus et al., 2005; Hobbie, 2003; Plaszczak and Wellner, 2005).

Although this new data intensive approach to research improves our understanding of ecosystems, challenges persist (Michener and Jones, 2012). Here, we propose a practical solution for one of the main challenges: how models and algorithms can be stored, documented, and managed in a way that allows their execution and interoperability.

Our premise is that most of the information regarding the design and implementation of ecological models or analytic methods is not available to the scientific community but rather is stored in individual scientists' digital or biological memories. Our thesis is that gathering together all this knowledge is critical to: a) expand our knowledge of the Earth as a system and advance our understanding of human impact on that system (Voinov and Cerco, 2010); and b) design and implement procedures for sustainable stewardship of natural resources in the Anthropocene era (Chapin et al., 2010; Crutzen, 2002). The creation of tools to preserve and manage algorithms, models, and scientific workflows will enhance code sharing and model reuse (Holzworth et al., 2010), and thereby help boost Ecology into taking its place as one of the so-called "big sciences", which, as their main features, encourage the growth of digital repositories, documentation of data and scientific processes, plus the creation of technical infrastructures to facilitate international collaboration (Borgman et al., 2007).

We argue that it is time for ecologists to face the challenge of storing and documenting their models and algorithms. Just as a few decades ago the need for primary data repositories was obvious; today the creation of model repositories should be considered the next step in that trend of improving data management. On the other hand, model documenting and code sharing is becoming indispensable in all sciences that depend on computation (Ince et al., 2012).

It will not be easy to create cyberinfrastructures to store, document, and execute ecological models, but Ecology as a science must confront this problem soon. Ecologists will be forced to address several issues that could change the way they do research: a) any model or analytical procedure should be documented using metadata standards. This is essential to promote both the automatic connectivity of any models as well as code reuse. b) Modelers should get used to uploading their code to model repositories. Despite the initial resistance, ecologists will recognize that model sharing improves integrated modeling and the reproducibility of analyses. c) The design and implementation of model repositories requires tackling a set of significant technical challenges, model coupling perhaps being the main one. Coupling the execution of several connected models requires a high degree of integration between model metadata, model data sources, timing of different models, model parameters, etc. In this sense, the process of creating model repositories in Ecology will be facilitated by the progressive awareness of scientists regarding model documenting, code sharing, and model reusing.

In this work, we propose a new conceptual framework to develop a tool to manage models for ecologists and environmental managers. The main interest of this conceptual framework (described in Section 3) is that it has been designed by combining the major advances made in other scientific areas (described in Section 2). We will also illustrate an implementation of this conceptual framework in a real model/workflow repository (Section 4). The result, called ModeleR (Pérez-Pérez et al., 2012) is the core of an information system that manages the data collected by the global change monitoring program of Sierra Nevada (Spain) LTER platform. We will show two case studies describing different types of models and workflows that can be documented and executed with ModeleR. We will demonstrate that using ModeleR is a realistic way to manage models and workflows in Ecology research groups. It can be considered a locally deployable tool that could help scientists to improve the way they document, store, share, and execute their models and workflows. This local approach makes ModeleR different from other tools used for these tasks.

2. Advances in model storing, documenting, and executing

The challenge of storing, documenting, and executing models and workflows is not unique to Ecology, but has also arisen in other areas of science and technology that have followed a similar path: from gathering and documenting primary data to creating complex models, and finally to designing and implementing tools to document and execute those models. Thus, Molecular Biology (Buckingham, 2007; Snoep et al., 2006), Earth System Science (Peckham et al., 2013), or even Process Systems Engineering (Kuntsche et al., 2011) have achieved noteworthy advances in model storing, documenting, and executing.

In this section, we describe these advances (Table 1 for a summary). This shared history among different disciplines could provide a valuable set of lessons that would be helpful to all of these fields. Below, we describe the most relevant ones regarding Molecular Biology and Earth Science.

Molecular Biology has advanced in the design and implementation of model repositories. With the immense scientific benefits gained by researchers after the creation of primary data repositories such as GeneBank [15], the documenting of algorithms and models has come to be considered obvious and necessary. According to Buckingham (Buckingham, 2007), model repositories should allow model documenting (by means of XML schemas or ontologies), must be connected to primary data repositories, and should be designed using the concept of web service. These concepts have been implemented in several tools (Snoep et al., 2006). We can highlight Biomodels (Li et al., 2010), a repository of peer-reviewed, curated, published, versionable, and parameterizable computational models. To promote the use and growth of Biomodels, some publishers encourage authors to upload their models here after publication. Other initiatives such as myExperiment (Goble et al., 2010) enable workflows to be shared among scientists, as in a social network. The major contributions made by Molecular Biology to the idea of a tool to manage ecological models include: a) the strong connection between primary data repositories and model repositories; and b) the efforts made to promote the use of model documenting and executing tools in the scientific community.

Earth System Science develops comprehensive and highly integrated models describing the interaction between atmosphere, hydrosphere, lithosphere, and biosphere. This need for integration has prompted the creation of tools similar to model repositories that: a) enable the coupling of model execution (Bulatewicz et al., 2009; Castronova et al., 2012); b) manage model versioning (Thornton et al., 2005); c) track computational provenance of models (Dozier and Frew, 2009; Frew et al., 2008); and d) create models collaboratively thanks to community modeling systems (Voinov et al., 2010). Major contributions of Earth System Science to the idea of a tool to manage ecological models have been the

Table 1
Summary of the notable progress achieved by different disciplines regarding model storing, documenting, and executing models and workflows.

Name	References	Scientific discipline	Key features
Biomodels	Li et al., 2010	Molecular biology	Connection with biological data sources. Web interface to search models using ontologies.
myExperiment	Goble et al., 2010	Molecular biology	Use of social networks to promote workflow and model sharing.
JWS Online Cellular Systems Modelling	Olivier and Snoep, 2004	Systems biology	Code repository able to interactively run kinetic models via web.
ECOBAS	Benz and Hoch, 2001	Ecology	Model documenting.
Meta-modeling	Keller and Dungan, 1999	Ecology	Framework helping to process model construction.
Simecol	Petzoldt, 2007	Ecology	R-package to implement, simulate, and share ecological models.
Meece (Marine ecosystem evolution in a changing environment)	http://www.meece.eu/Library.aspx	Marine ecology	Catalog of documented models.
OpenMI	Turuncoglu et al., 2013	Hydrology, Earth Sciences	Software component interface to linking any combination of models, databases, and analytical and visualization tools.
ESMF (Earth System Modeling Framework)	Hill et al., 2004	Earth Sciences	Software infrastructure to promote connectivity between models.
CSDMS (Community Surface Dynamics Modelling System)	Peckham et al., 2013; Voinov et al., 2010	Earth Sciences	Component-based system to promote connectivity among models.
MOSAIC	Kuntsche et al., 2011	Process systems engineering	Documenting equations using a systematic approach.
CoMSES (The Network for Computational Modeling for SocioEcological Science)	http://www.openabm.org/models	Social and life sciences	Library containing descriptions of agent-based models.

advances in model execution and system modularity. These challenging issues are especially important due to the complexity in the design and implementation of standards to document and couple model executions. Notable among the different initiatives that are trying to design and implement tools to manage models (document, couple executions, share code, etc.) is CSDMS (Community Surface Dynamics Modeling System) (Overeem et al., 2013; Peckham et al., 2013), an NSF (National Science Foundation) program with the aim of integrating a wide variety of Earth surface processes considering different temporal and spatial scales. CSDMS also contains a source code repository where modelers can upload or download models created by themselves or by other scientists. Each model can be documented in detail, describing the processes involved, input and output datasets, bibliographic references, and technical specifications. On the other hand, OpenMI (Goodall et al., 2013; Gregersen et al., 2007) provides a standard interface to facilitate the data exchange between any models.

3. Conceptual framework for an envisioned tool to store, document and execute both models and workflows

Our proposed conceptual framework combines certain major advances (Table 1) made by the above-mentioned sciences regarding the issue of model repositories. In this section, we describe this framework according to its main functions: model documenting, connectivity, model execution, and model versioning. Fig. 1 shows a graphic description of this framework.

3.1. Model documenting

Model documenting could be considered the first and the most important function of a model/workflow management tool. Our proposed documenting schema is based on ECOBAS (Benz and Hoch, 2001; Strube et al., 2008), as we are considering a hierarchical documenting schema. A modeler can document a model using three levels of detail: a) general model description (aims, methods, bibliography, authors, links); b) general information included in the first level plus detailed technical information and some basic information about the mathematics of the model; c) Information included in the second level plus detailed information about the mathematics of the model (algorithm description, parameters, connection with source data, execution rules). This documenting structure should be versatile enough to cope with both statistical and analytical models, and others that like ABMs (agent based models) have their own documenting protocols (ODD: overview, design concepts, and details (Grimm et al., 2010; Müller et al., 2013)). This documenting framework also allows the documenting of models that contain only one or several algorithms not implemented.

The "depth" of the documentation process depends on the modeler's knowledge of how to use the model. The most detailed level of documenting should also include the description of input and output datasets as an inherent part of a model documenting process. Given that input/output datasets are extraordinarily diverse in environmental sciences (Reichman et al., 2011), the repository should be able to "speak several metadata dialects" (Nogueras-Iso et al., 2004) used to document different types of datasets. This could be by the creation of crosswalks between metadata standards. Thus, the proposed tool could read metadata catalogs written with different specifications and add these datasets as input data.

3.2. Connectivity among models

The capability of linking models with data sources and outputs provides a powerful framework to promote connectivity between different models stored in a repository. We are proposing that when the outcome of a model (e.g. M1) can be used as an input dataset by another model (e.g. M2), then both models become connected (Lundström et al., 2007; Zerger and Wealands, 2004). This connection becomes real when models are executed. Execution of M2 requires the previous execution of M1, for example. A temporal coupling of executions is also needed, as the outcome for M1 in the first time step is used by M2 for its first step execution, too. This framework allows for the existence of intermediate models that calibrate output datasets to make them useful for other models.

Linking models relies on web services (Goble et al., 2010; Goodall et al., 2013). These services would be interoperable using metadata specifications and would supply information about input/output datasets and algorithm characteristics. There would be a web service for each model and another one for the input/output dataset.

Linking models that are stored in different model repositories could yield the creation of a federation of model repositories. This federation of repositories, based on the statement that information

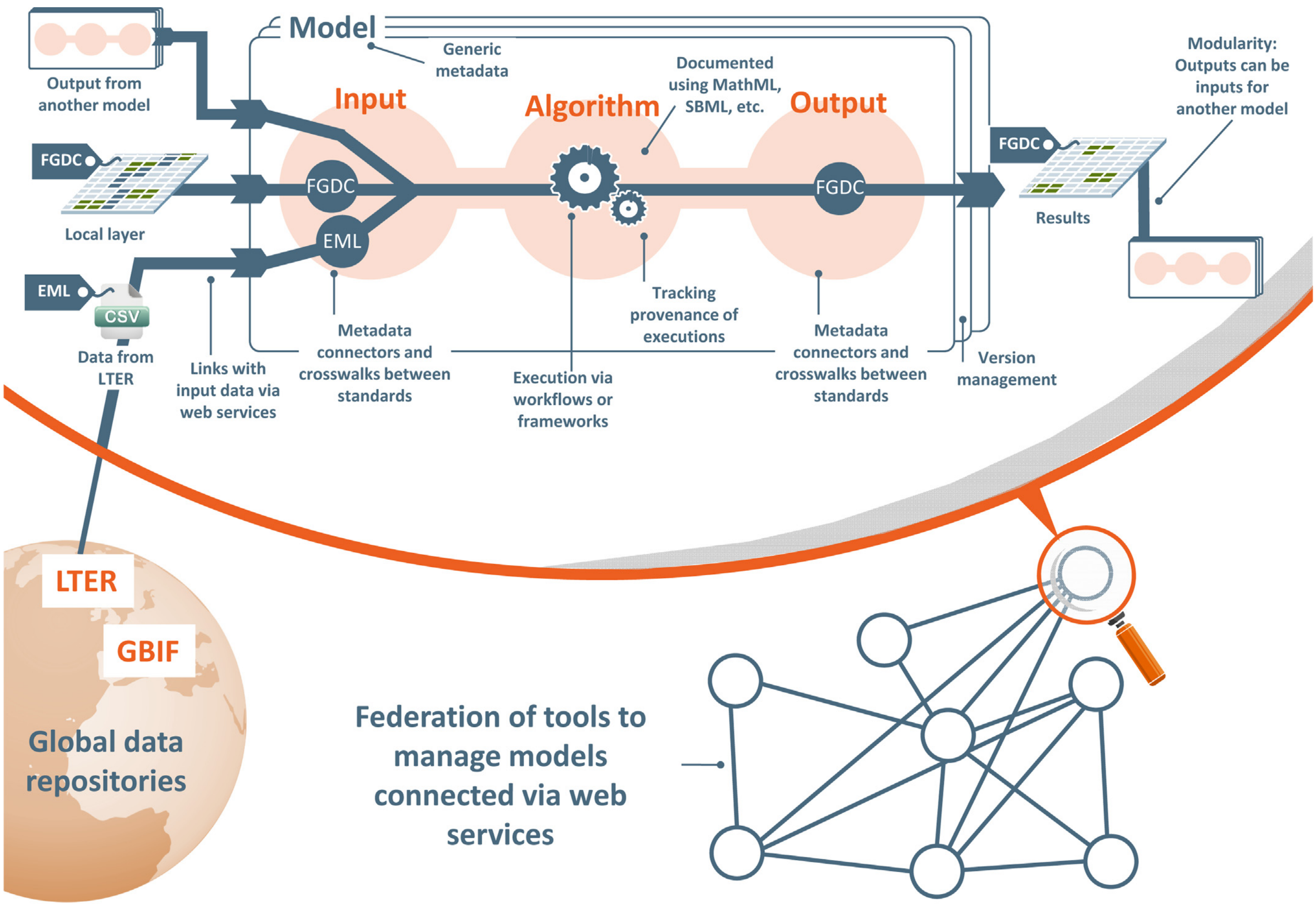

Fig. 1. Scheme showing the most important functions of the proposed tool. The envisioned infrastructure should be able to describe diverse ecological models, following a gradient of levels of documentation: from simply adding a name and a unique identifier (Bafna et al., 2008; Wang, 2007) to providing a detailed description of the mathematical algorithms used by the model (Hucka et al., 2003; Miner, 2005). Input/output datasets should also be documented using metadata specifications such as Ecological Metadata Language (Michener, 2006) and others. Finally, models documented in great detail could be executed by the system. This conceptual framework takes into account the ability for a federation of model repositories. Local repositories could connect to some others via web services.

must be stored and curated wherever it is created, is best suited to promote synergy among scientists and also to meet their demands of managing the models that they create. Web-based simulation tools and component-based programming could be useful to implement the idea of a federation of local repositories (Byrne et al., 2010).

3.3. *Model execution*

Only models with the most detailed level of documenting could be executed by the system. When executing a model, it is important to track specific actions by algorithms in input datasets. This feature, called 'tracking provenance' (Dozier and Frew, 2009; Frew et al., 2008), allows the optimization of model execution and should be included in any framework aiming to execute ecological models. The proposed web tool needs to be able to execute both workflows, such as Kepler (Altintas et al., 2004) or Taverna (Oinn et al., 2004), and scriptable code. The output (only those considered definitive or useful for other models) should be properly documented, too, in order to allow model connectivity.

3.4. *Model versioning*

Besides the above-described functions, a tool to manage ecological models and workflows should consider the changes that algorithms as well as input datasets undergo during the lifecycle of a model. The versioning system should include both the objects to be versioned (algorithms, datasets, parameters) and a procedure to identify versions (Ogasawara et al., 2009). A model versioning system helps to (Altmanninger et al., 2009): a) maintain the history of the evolution of models, b) support collaboration between modelers, and c) manage different development branches in a given model. According to Ogasawara (Ogasawara et al., 2009) we are proposing a versioning system where each versioned object in any model (algorithm, input, and output) has an object identifier (OID). This means that any change in input datasets, model algorithm or model parameters would imply the creation of a new OID. There are different types of OIDs: digital object identifier (Parr et al., 2012; Wang, 2007) is an international standard for identifying digital objects. LSID (life-science identifiers; (Bafna et al., 2008) can be used to name and locate information on the web.

4. ModeleR: a web based model and workflow management system

Besides formulating a conceptual framework, we have also implemented some of its main functions within a functional tool, called ModeleR (Pérez-Pérez et al., 2012). This web application has been built in the context of the Sierra Nevada (S Spain) global change monitoring program. Sierra Nevada is a Biosphere reserve, a

National Park, and also an LTER (long term ecological research) platform.

We have implemented in ModeleR the documenting functionality previously described in the conceptual framework. Thus, ModeleR can document different types of ecological models. It also allows the connection of a given model to a dataset documented using EML (ecological metadata language; Fegraus et al., 2005). If the model is documented at the most detailed level, the system automatically creates an initial prototype of a Kepler workflow (Altintas et al., 2004). When the user completes this prototype (adding the code needed, all remaining actors and parameters), it can be uploaded to ModeleR and then the model can be executed by the system. Thus, these workflows are used to guide different execution steps within one model. In addition, if the model is not a Kepler workflow, the system accepts any kind of code to execute a model. The only requisites are to install the execution engine (must be scriptable; e.g. R, MatLab, or ESRI scripts) in the server where ModeleR is running and upload the code. This requisite could be a limitation if those scripts are created using commercial software, due to the cost associated with such software. Thus, using commercial software causes dependence and limitations in the system. We have also implemented the connectivity function previously described: model outputs could be used by another model as inputs. To promote the idea of a federation of model repositories, we have created ModeleR using a service-oriented architecture (SOA), which would allow interoperability and connectivity with other similar tools. Finally, model creation and documentation can be undertaken collaboratively thanks to a blogging system associated to ModeleR.

5. ModeleR in action: two study cases in the context of a global change monitoring program

Currently, we are using ModeleR to document and execute models and analytical processes associated with the Sierra Nevada LTER platform. This model management tool, which can be accessed via web (http://modeler.obsnev.es/), has become the core of the platform's information system. The global change monitoring program under way in Sierra Nevada offers a great array of datasets and algorithms that have been used to test both ModeleR as well as the conceptual framework that supports it. This monitoring program provides information about more than 100 environmental variables belonging to 48 monitoring protocols. This great amount of information is processed via ModeleR to gain useful knowledge to assist environmental decision making.

ModeleR can document several types of models from aggregation/summarization workflows (e.g. processing data from meteorological stations) to species distribution models (Jetz et al., 2012) (Fig. 2) or dynamic models such as agent based models (Grimm et al., 2005) (Fig. 3). These two examples illustrate how ModeleR works and how models can be connected to each other by documenting their inputs/outputs.

6. Concluding remarks

The creation of tools able to document, store, and execute ecological models collaboratively (Byrne et al., 2010) will help advance Ecology in two ways: by bolstering the capacity of creating relevant knowledge and by improving its capacity to transfer that knowledge to decision makers.

Model documentation is a beneficial practice in modeling (Scholten, 2008), facilitating the process of conceptualization, mathematical description, and real implementation of a given model (Keller and Dungan, 1999). A tool to document and execute workflows/models will also improve model outreach among different scientists. Model sharing should also allow modelers to advance in the reproducibility of their analyses (Cassey, 2006). Model sharing and code reuse will strengthen the environmental management decisions that could be based upon those models (Scholten, 2008).

Despite the notable advantages provided by model documentation and model repositories, these tools are not yet very successful in Ecology. Some scientists argue that using these tools is excessively time consuming. This might be disputed, since, according to our calculation, the time to document a complex model (such as the examples provided above) is less than two hours using ModeleR. This process is straightforward if beforehand the user has locally described and documented the intermediate steps given to build a complex model. In other words, the better the local process of documenting a model, the easier it will be to describe it in any model/workflow management tool.

Another reason that could explain the lack of success of model management tools in Ecology is the meager reward that scientists receive on documenting and uploading models into repositories. Molecular Biology is implementing a strategy to properly reward scientists that document their models. Some publishers encourage authors to upload their models into repositories after publishing them in journals (Li et al., 2010). We believe that a similar strategy applied to ecological models could promote the use of these tools.

Regarding model sharing via repositories, it is important to examine the issue of licensing and model sharing policy. Another drawback in implementing model repositories could be that modelers might prefer to preserve their intellectual rights. On the other hand, model repositories are most useful when the concept of "open access" is applied. That is, the higher the rate of sharing models, the greater the synergy generated. To cope with both scenarios, model repositories must implement sharing policies. Although this issue has not been fully addressed in this work, we propose an approach similar to the one created by GBIF (Global Biodiversity Information Facility), which has built a data portal with almost 400 million data records on worldwide species distribution. Part of their success is due to one of the statements described in their data sharing policy: "Data providers retain all rights and responsibilities associated with the data they make available". Data providers also decide whether they provide whole datasets showing species occurrence or only metadata. We propose to transfer this approach to model documenting.

We are proposing a tool capable of documenting and executing several types of ecological models or analytical processes. The tool described here will help to extend the paradigm of metadata from datasets to models and algorithms. Finally, our approach is also taking into account the scalability needs of individuals, research groups, and institutes. The overarching idea is to create an international federation of tools to manage models/workflows with a high degree of connectivity. The benefits of this tool can be summed up with a single word: synergy. The easier it is to share how we analyze ecological data to get useful knowledge, the stronger the collaboration will be among peers.

Acknowledgments

This work has been done in the collaborative framework of the "Sierra Nevada Global Change Observatory" Project from the Environment Department of Andalusian Regional Government, the Sierra Nevada National Park and the Spanish Biodiversity Foundation ("Fundación Biodiversidad"). Funding was provided by the project, "Creación de un repositorio de modelos para la red de información ambiental de Andalucía", and MIGRAME (RNM 6734) both from the Andalusian Regional Government. We thank David Nesbitt for English review.

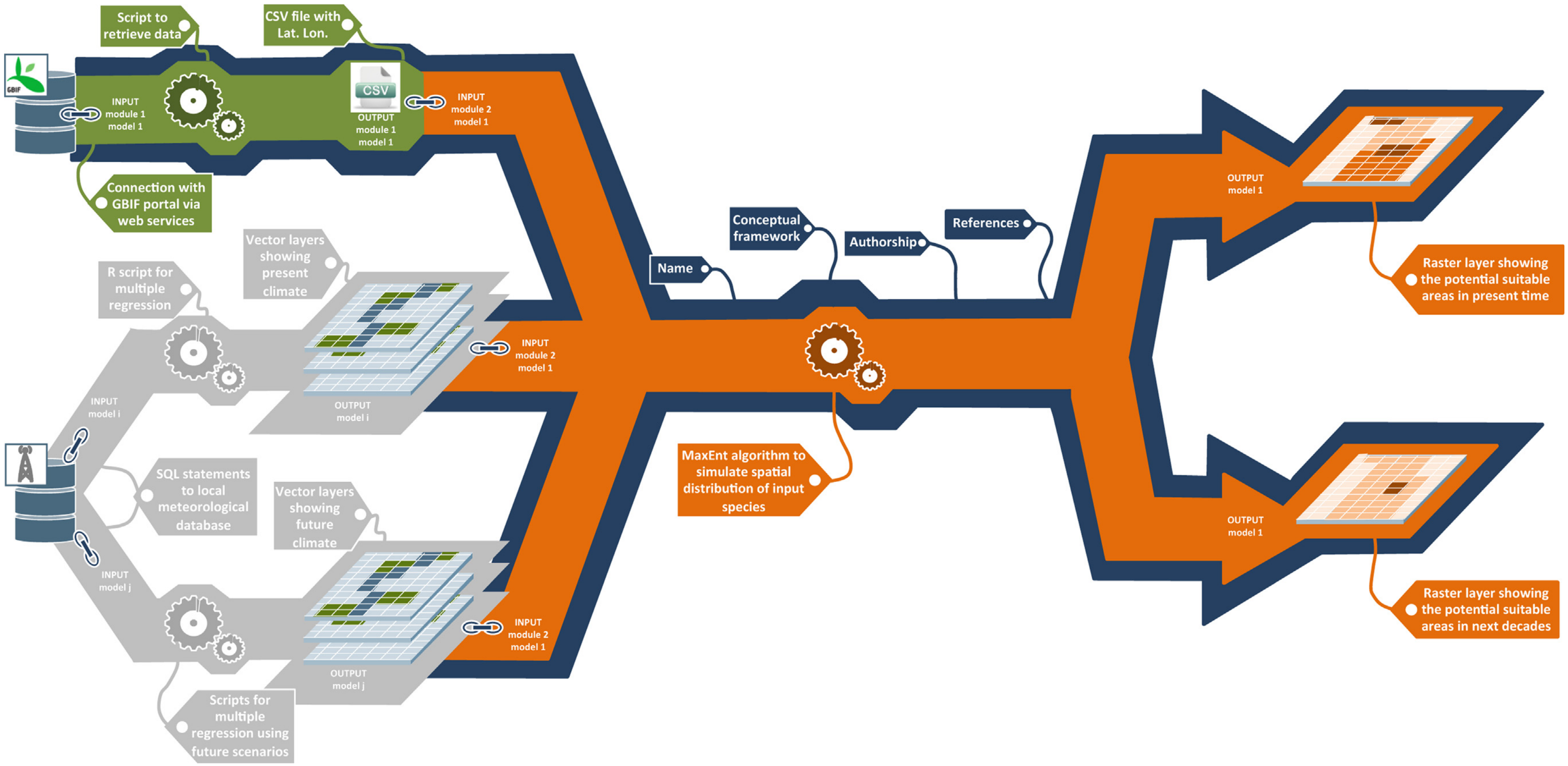

Fig. 2. This diagram shows how a species distribution model (SDM) is documented using ModeleR. SDMs are workflow based models that aim to provide a spatial representation of the area potentially used by a particular species according to the distribution of certain environmental variables. The whole model is described using the first level of documenting (dark-blue labels). In a more detailed view, the model can be described as two different modules connected via input/output elements. The first module (green) is in charge of retrieving presence data from GBIF (Global Biodiversity Information Facility) using web services. This dataset represents the presence distribution of different species in the landscape. The second module (orange) uses as input data the outputs created by the previous module and two outputs created by different models (Model 2, present climate; Model 3, future climate). These last two datasets are the environmental variables that are assumed to influence species distribution. The SDM uses a maximum entropy algorithm to simulate the potential distribution of the target species. This algorithm is also able to project the “potential habitat” over the present and future environmental variables, yielding two output layers: one showing the areas suitable for the target species at present and another one showing the suitability in the future. In this case, all algorithms are implemented in R or different Kepler’s actors. This link (http://goo.gl/vhzgM) shows a summary of the documentation process for this model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

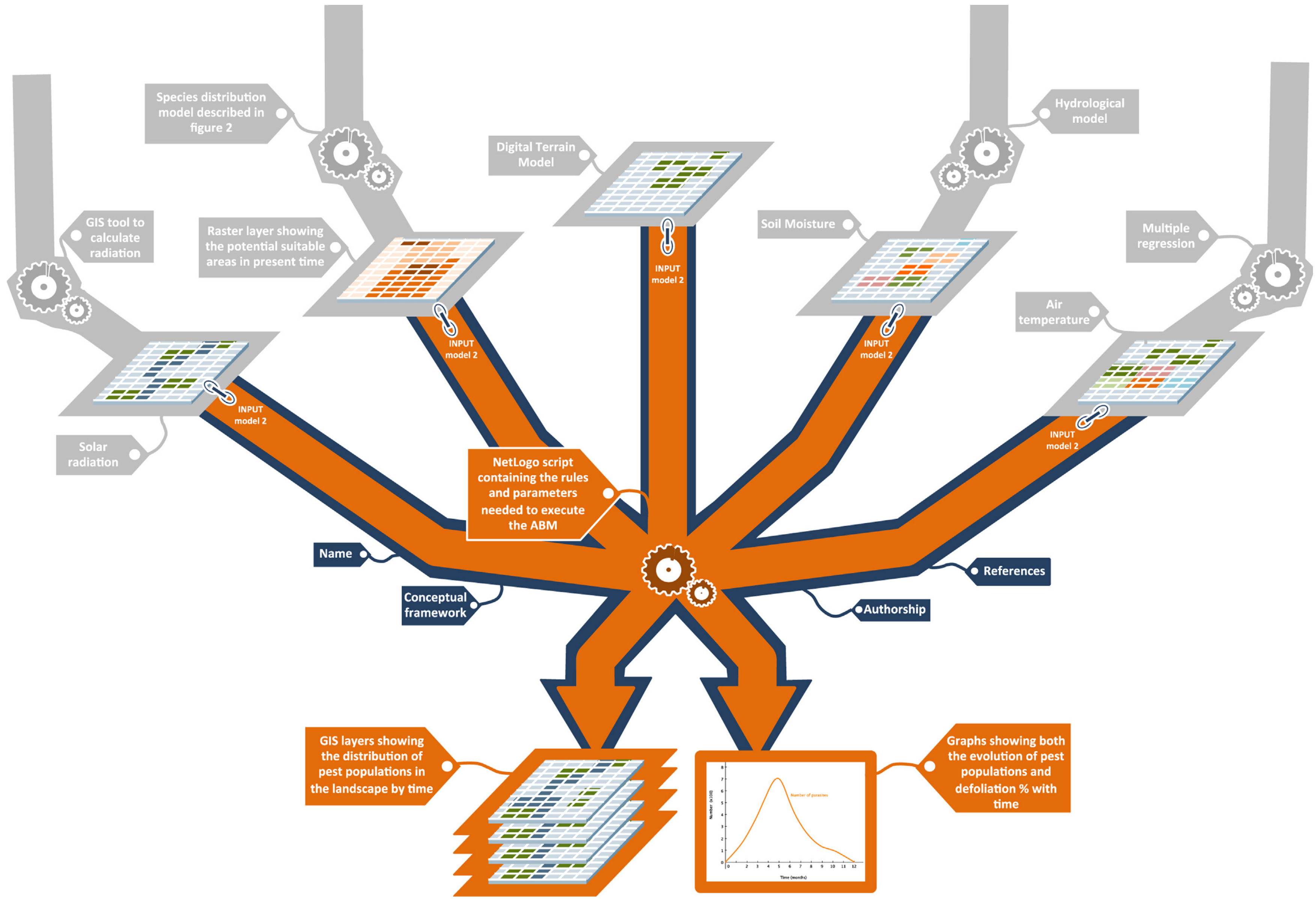

Fig. 3. This figure shows how an agent based model (ABM) is documented using ModeleR. In this specific case, ABMs can be considered dynamic models that simulate the behavior of a set of agents that act in a given landscape. This example describes the dynamics between a forest pest and its host tree. The input data are layers describing the status of some environmental variables (solar radiation, areas potentially suitable for host species, terrain, or soil moisture) that are important for an agent's activities. Here, we have considered two types of model output: a) GIS layers showing the spatial distribution of hosts and pest agents in the landscape and b) graphs showing both the evolution of pest numbers and the defoliation percentage of hosts with time in the overall landscape. The model is described using the first level of documentation (dark-blue labels). The next level of documentation involves the description of a single module and its inputs/outputs. Some inputs are the outputs created by other models implemented in ModeleR. ABM's core is a script containing both the behavior of each agent type and the relationships of these agents with the landscape. This script can be coded with NetLogo or any other modeling software. The most detailed level of documenting implies uploading the NetLogo script. Users can also include all the parameters needed by the algorithm to be executed. This link (http://goo.gl/DDJw6) shows a summary of the documentation process for this model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

References

Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., Mock, S., 2004. In: Kepler: an Extensible System for Design and Execution of Scientific Workflows, Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004. IEEE.

Altmanninger, K., Seidl, M., Wimmer, M., 2009. A survey on model versioning approaches. Int. J. Web Inf. Syst. 5, 271–304.

Bafna, S., Humphries, J., Miranker, D.P., 2008. Schema driven assignment and implementation of life science identifiers (LSIDs). J. Biomed. Informa. 41, 730–738.

Benz, J., Hoch, R., 2001. ECOBAS — modelling and documentation. Ecol. Model. 138, 3–15.

Borgman, C.L., Wallis, J.C., Enyedy, N., 2007. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. Int. J. Digit. Libr. 7, 17–30.

Buckingham, S., 2007. To build a better model. Nat. Methods 4, 367–373.

Bulatewicz, T., Yang, X., Peterson, J.M., Staggenborg, S., Welch, S.M., Steward, D.R., 2009. Accessible integration of agriculture, groundwater, and economic models using the Open Modeling Interface (OpenMI): methodology and initial results. Hydrol. Earth Syst. Sci. Discuss. 6, 7213–7246.

Byrne, J., Heavey, C., Byrne, P.J., 2010. A review of web-based simulation and supporting tools. Simul. Model. Pract. Theory 18, 253–276.

Cassey, P., 2006. Reproducibility and repeatability in ecology. BioScience 56, 958.

Castronova, A.M., Goodall, J.L., Ercan, M.B., 2012. Integrated modeling within a hydrologic information system: an OpenMI based approach. Environ. Model. Softw., 1–11.

Chapin, F.S., Carpenter, S.R., Kofinas, G.P., Folke, C., Abel, N., Clark, W.C., Olsson, P., Smith, D.M.S., Walker, B., Young, O.R., Berkes, F., Biggs, R., Grove, J.M., Naylor, R.L., Pinkerton, E., Steffen, W., Swanson, F.J., 2010. Ecosystem stewardship: sustainability strategies for a rapidly changing planet. Trends Ecol. Evol. 25, 241–249.

Crowley, P.H., 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. Annu. Rev. Ecol. Syst. 23, 405–447.

Crutzen, P., 2002. Geology of mankind. Nature 415, 2002.

Dozier, J., Frew, J., 2009. Computational provenance in hydrologic science: a snow mapping example. Philos. trans. Ser. A Math. Phys. Eng. Sci. 367, 1021–1033.

Fegraus, E., Andelman, S., Jones, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86, 158–168.

Frew, J., Metzger, D., Slaughter, P., 2008. Automatic capture and reconstruction of computational provenance. Concurr. Comput. Pract. Exp. 20, 485–496.

Goble, C.A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., De Roure, D., 2010. myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Res. 38, W677–W682.

Goodall, J.L., Saint, K.D., Ercan, M.B., Briley, L.J., Murphy, S., You, H., DeLuca, C., Rood, R.B., 2013. Coupling climate and hydrological models: interoperability through web services. Environ. Model. Softw. 46, 250–259.

Green, J.L., Hastings, A., Arzberger, P., Ayala, F.J., Cottingham, K.L., Cuddington, K., Davis, F., Dunne, J. a, Fortin, M.-J., Gerber, L., Neubert, M., 2005. Complexity in ecology and conservation: mathematical, statistical, and computational challenges. BioScience 55, 501.

Gregersen, J.B., Gijsbers, P.J. a, Westen, S.J.P., 2007. OpenMI: Open modelling interface. J. Hydroinforma. 9, 175.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F., 2010. The ODD protocol: a review and first update. Ecol. Model. 221, 2760–2768.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991 (New York, N.Y.).

Hill, C., DeLuca, C., Suarez, M., 2004. The architecture of the earth system modeling framework. Comput. Sci. Eng. 6, 18–28.

Hobbie, J.E., 2003. Scientific accomplishments of the long term ecological research Program;: an introduction. BioScience 53, 17–20.

Holzworth, D.P., Huth, N.I., de Voil, P.G., 2010. Simplifying environmental model reuse. Environ. Model. Softw. 25, 269–275.

Hucka, M., Finney, a, Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, a. P., Bornstein, B.J., Bray, D., Cornish-Bowden, a, Cuellar, a. a, Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, a, Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531.

Ince, D.C., Hatton, L., Graham-Cumming, J., 2012. The case for open computer programs. Nature 482, 485–488.

Jetz, W., McPherson, J.M., Guralnick, R.P., 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends Ecol. Evol. 27, 151–159.

Jones, M.B., Schildhauer, M.P., Reichman, O.J., Bowers, S., 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annu. Rev. Ecol. Evol. Syst. 37, 519–544.

Keller, R.M., Dungan, J.L., 1999. Meta-modeling: a knowledge-based approach to facilitating process model construction and reuse. Ecol. Model. 119, 89–116.

Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., Hooker, G., 2009. Data-intensive science: a new paradigm for biodiversity studies. BioScience 59, 613–620.

Kuntsche, S., Barz, T., Kraus, R., Arellano-Garcia, H., Wozny, G., 2011. MOSAIC a web-based modeling environment for code generation. Comput. Chem. Eng. 35, 2257–2273.

Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Le Novère, N., Laibe, C., 2010. BioModels database: an enhanced, curated and annotated resource for published quantitative kinetic models. BMC Syst. Biol. 4, 92.

Lundström, C., Kytzia, S., Walz, A., Gret-Regamey, A., Bebi, P., 2007. Linking models of land use, resources, and economy to simulate the development of mountain regions (ALPSCAPE). Environ. Manag. 40, 379–393.

Metzger, K.J., Klaper, R., Thomas, M.a., 2011. Implications of informatics approaches in ecological research. Ecol. Inf. 6 (1), 4–12.

Michener, W.K., 2006. Meta-information concepts for ecological data management. Ecol. Inf. 1, 3–7.

Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27, 85–93.

Miner, R., 2005. The importance of MathML to communication. Not. Am. Math. Soc. 52, 532–538.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., Schwarz, N., 2013. Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. Environ. Model. Softw. 48, 37–48.

Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R., Muro-Medrano, P.R., 2004. Metadata standard interoperability: application in the geographic information domain. Comput. Environ. Urban Syst. 28, 611–634.

Ogasawara, E., Rangel, P., Murta, L., Werner, C., Mattoso, M., 2009. Comparison and versioning of scientific workflows. In: 2009 ICSE Workshop on Comparison and Versioning of Software Models, pp. 25–30.

Oinn, T., Addis, M., Ferris, J., Marvin, D., Carver, T., Pocock, M.R., Wipat, A., 2004. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20, 2004.

Olivier, B.G., Snoep, J.L., 2004. Web-based kinetic modelling using JWS Online. Bioinformatics 20, 2143–2144.

Overeem, I., Berlin, M.M., Syvitski, J.P.M., 2013. Strategies for integrated modeling: the community surface dynamics modeling system example. Environ. Model. Softw. 39, 314–321.

Parr, C.S., Guralnick, R., Cellinese, N., Page, R.D.M., 2012. Evolutionary informatics: unifying knowledge about the diversity of life. Trends Ecol. Evol. 27, 94–103.

Peckham, S., Hutton, E., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: the design of CSDMS. Comput. Geosci. 53, 3–12.

Pérez-Pérez, R., Benito, B.M., Bonet, F.J., 2012. ModeleR: an environmental model repository as knowledge base for experts. Expert Syst. Appl. 1, 1829–1841.

Petzoldt, T., 2007. simecol: an Object-Oriented Framework for Ecological Modeling in R. J. Stat. Softw. 22.

Plaszczak, Pawel, Wellner Jr., R., 2005. Grid Computing the Savvy Manager's Guide. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331, 703–705 (New York, N.Y.).

Scholten, H., 2008. Good modelling practice. In: 13th JISR-IIASA Workshop on Methodologies and Tools, pp. 15–31.

Snoep, J.L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V., 2006. Towards building the silicon cell: a modular approach. Biosystems 83, 207–216.

Strube, T., Benz, J., Kardaetz, S., Brüggemann, R., 2008. ECOBAS — A tool to develop ecosystem models exemplified by the shallow lake model EMMO. Ecol. Inf. 3, 154–169.

Thornton, P., Cook, R., Braswell, B., 2005. Archiving numerical models of biogeochemical dynamics. EOS 86, 6–7.

Turuncoglu, U.U., Dalfes, N., Murphy, S., DeLuca, C., 2013. Toward self-describing and workflow integrated Earth system models: a coupled atmosphere-ocean modeling system application. Environ. Model. Softw. 39, 247–262.

Voinov, A., Cerco, C., 2010. Model integration and the role of data. Environ. Model. Softw. 25, 965–969.

Voinov, A., DeLuca, C., Hood, R., Peckham, S., 2010. A community approach to Earth systems modeling. Eos, 91.

Wang, J., 2007. Digital object identifiers and their use in libraries. Ser. Rev. 33, 161–164.

Whitlock, M.C., 2011. Data archiving in ecology and evolution: best practices. Trends Ecol. Evol. 26, 61–65.

Zerger, A., Wealands, S., 2004. Beyond modelling: linking models with GIS for flood risk management. Nat. Hazards 33, 191–208.