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

A R T I C L E I N F O

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

Keywords:
Algorithms
Metadata
Model
Documenting
Workflow

A B S T R A C T

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