Environmental Modelling & Software 87 (2017) 49–63

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

An overview of the model integration process: From pre-integration assessment to testing

Getachew F. Belete [a, *], Alexey Voinov [a], Gerard F. Laniak [b]

[a] University of Twente, ITC, 7500 AE Enschede, Netherlands
[b] US Environmental Protection Agency, Office of Research and Development, USA

A R T I C L E I N F O

Article history:
Received 15 March 2016
Received in revised form
30 August 2016
Accepted 30 October 2016

Keywords:
Integrated modeling
Interoperability
Interfaces
Wrapping
Components
Web services

A B S T R A C T

Integration of models requires linking models which can be developed using different tools, methodologies, and assumptions. We performed a literature review with the aim of improving our understanding of model integration process, and also presenting better strategies for building integrated modeling systems. We identified five different phases to characterize integration process: pre-integration assessment, preparation of models for integration, orchestration of models during simulation, data interoperability, and testing. Commonly, there is little reuse of existing frameworks beyond the development teams and not much sharing of science components across frameworks. We believe this must change to enable researchers and assessors to form complex workflows that leverage the current environmental science available. In this paper, we characterize the model integration process and compare integration practices of different groups. We highlight key strategies, features, standards, and practices that can be employed by developers to increase reuse and interoperability of science software components and systems.

Contents

1. Introduction 50
2. Pre-integration assessment 51
3. Technical integration of models: preparing models for integration and orchestration of participating models during simulation 53
  3.1. Preparing models for integration 53
  3.2. Orchestration of participating models during simulation 54
  3.3. Commonly used approaches for technical integration of models 54
    3.3.1. Component-based approach for technical integration of models 54
    3.3.2. Web services and service-oriented approach for technical integration of models 55
    3.3.3. Hybrid of component-based and web service-based approach for technical integration 56
    3.3.4. Tailor-made (custom-made) techniques for technical integration of models 56
  3.4. Interface standards and interoperability 56
  3.5. Performance optimization and related issues in orchestrating simulations 57
4. Data interoperability 57
  4.1. Issues on data interoperability 57
  4.2. Hard-coding for data interoperability 58
  4.3. Framework specific annotations for data interoperability 58
  4.4. Controlled vocabulary and ontology for data interoperability 58
5. Testing integration 59
6. Discovery, accessibility and ease of use of integrated systems 60
7. Discussion and conclusions 60

* Corresponding author.
E-mail addresses: getfeleke@gmail.com (G.F. Belete), aavoinov@gmail.com (A. Voinov), laniak.gerry@epa.gov (G.F. Laniak).

http://dx.doi.org/10.1016/j.envsoft.2016.10.013
1364-8152/

8. Recommendations ... 61
8.1. User interface ... 61
8.2. Model integration ... 62
8.3. Science components ... 62
Disclaimer ... 62
Acknowledgment ... 62
References ... 62

1. Introduction

A model is a simplified abstraction of reality (Schmolke et al., 2010; Voinov, 2008). To answer questions related to environmental problems, scientists and policy-makers, may need to address complex issues at regional, continental, and even global scales (Harris, 2002; Verburg et al., 2008). Developing all-inclusive models is difficult, however, and we must integrate individual models that represent specific domains (Gregersen et al., 2007). Integration of models is becoming more important due to an increased desire to understand, investigate, and mitigate human impact on the environment (Laniak et al., 2013; Voinov and Cerco, 2010). The challenge is making standalone models discoverable, reusable, and interoperable (Goodall et al., 2011). Here, discoverability means the availability of meta information of a resource so it can be effectively searched for, located and understood to ensure proper use (Erl, T., 2008b). Reusability is how existing software can, in whole or in part, be used to develop new software (Frakes and Kang, 2005), as well as the degree to which an asset can be used in more than one system or in building other assets (ISO/IEC, 2011). Reuse also implies a decision process that includes evaluating software for its ability to serve a new purpose. And, finally, interoperability is the ability of a system or a product to work with other systems or products without special effort by the user (Chang and Lee, 2004).

In this paper a component may represent an elemental physical/chemical/biological/social process (e.g., infiltration or advection) or a logical integration of elemental processes (e.g., a watershed model). A system may be a model (an integrated set of elemental processes) or modeling system. Note that a model can refer to a component or a system. Finally, when we refer to an integrated modeling system we are including both the collection of science components and the framework software that facilitates their orchestration. These definitions apply throughout the paper.

Data exchange and data manipulation are fundamental to integrated modeling systems (Argent, 2004; Leimbach and Jaeger, 2005), and are often constrained by technical and conceptual challenges. Technically, individual models are designed to serve as stand-alone components that serve unique purposes and goals. Conceptual challenges include resolving the different ways modelers and science domains represent data and knowledge. Interoperability of models can also be provided at different levels. At the technical level, models should be able to 'talk to each other' which requires automating data exchange, making models jointly executable, and ensuring repeatability and reproducibility of model chain configuration and processing (Knapen et al., 2013). At the semantic level, models should 'understand each other' by identifying and, if possible, bridging semantic differences in an automated manner. After semantic mediation, the dataset should be syntactically interoperable, i.e., data produced by one model are converted and formatted to serve as input for a receiving model.

We performed a literature review to improve our understanding of the model integration process and also to collect and present better strategies for building integrated modeling systems. At first, our review was limited to integrated climate change mitigation analysis systems. Due to the commonality of integration techniques with other environmental domains, however, we broadened the scope to include integration of models in general. Our search used key phrases of 'integration of models', 'integrated assessment study' and 'integrated modeling frameworks,' and we considered 38 articles (see Table 1), of which 18 focused on specific integration modeling systems (aka frameworks).

To provide structure and context for the paper, we look to model development process which includes requirements analysis, design, implementation, and testing phases (David et al., 2013; Moore and Tindall, 2005; Whelan et al., 2014), with these steps conducted in an iterative manner (Grimm et al., 2014; Jakeman et al., 2006). Model integration process follows a similar sequence and is also iterative (Holzworth and Huth, 2011; Holzworth et al., 2014), with a small portion of integration realized first, and more complex issues and functionalities incorporated through iterations. Keeping all this in mind, we divided model integration process into phases of pre-integration assessment, preparation of models for integration, orchestration of participating models during simulation, data interoperability, and testing (Fig. 1).

This process forms the basis for our review. As we conducted the review, it became clear that three topics deserve special attention due to their significant role: interface standards for interoperability (within the preparation of models for integration phase); performance in integrating models (within the orchestration phase); and discovery, accessibility, and ease of use which are process-wide considerations. A summary of this structure, along with brief descriptions of categories of methods employed, examples of existing modeling systems, and references are presented in Table 1.

We note that:

- Most papers emphasize only certain phases of model integration.
- Comparing all selected frameworks with all aspects of the integration process is challenging due to limitations of available information. We believe, however, that, collectively, the full range of methods is represented.
- The set of model integration frameworks in Table 1 represents a robust, albeit incomplete, representation of existing frameworks.

Findings related to phases of integration: pre-integration assessment, preparation of models for integration, model orchestration, data interoperability, and testing are described in Sections 2 through 5. To avoid repetition of content, we treat preparation of models for integration and orchestration of participating models during simulation together in Section 3. Discoverability, accessibility, and ease of use of integrated systems are addressed in Section 6. Section 7 presents discussion as well as conclusions, and finally recommendations for increasing efficiency and effectiveness of the model integration process are presented in Section 8.