SUNPRISM: An Approach and Software Tools for Collaborative Climate Change Research

Sohei Okamoto¹
Roger V. Hoang¹
¹College of Engineering
University of Nevada, Reno, USA
{okamoto, hoangr, dascalus, fredh}@cse.unr.edu

Sergiu M. Dascalu¹, Frederick C. Harris, Jr.¹
Noureddine Belkhatir²
²LIG Computer Science Laboratory
University of Grenoble, France
noureddine.belkhatir@iut2.upmf-grenoble.fr

Abstract— In order to simulate scientific models in various conditions climate change research involves using large sets of heterogeneous data. In many cases, for model interoperability a dataset from a given model or source needs to be transformed into a different data format required by a subsequent model or processor. This paper presents the SUNPRISM approach and software tools aimed at supporting collaborative scientific exploration via new capabilities for combining data transformations, model simulations, and output visualizations in application scenarios developed for climate change research. The SUNPRISM framework’s defining characteristics are a visual object-based interface for scenario configuration, a workflow-based environment that allows code generation and dataflow scenario execution, and data visualization capabilities for 3D environments, including for immersive virtual environments such as CAVE. The paper describes the proposed approach and its supporting tools, the SUNPRISM Scenario Manager and the SUNPRISM Visualizer, and illustrates them on an application scenario in which data from the National Digital Forecast Database is retrieved and visualized in 3D. A brief comparison with related work and an outline of directions for future work are also included in the paper.

Keywords-SUNPRISM approach; workflow-based environment; model interoperability; software tools; data visualization; collaborative research

I. INTRODUCTION

In 2008, a Nevada-wide collaborative project has been funded by NSF EPSCoR to focus on regional and sub-regional scale effects on climate change and ecosystem resources and support infrastructure developments for scientific research, education, outreach, and decision making [18]. Part of the project, which will run until 2013, new datasets are being created to understand historical climate change processes and evolution, create model outcome views, and simulate scientific models in various conditions. Data types range from climate variables such as temperature, pressure, wind speed, and radiation, and could be aggregated into a large variety of different datasets [12, 18, 19].

Unfortunately, available datasets are not efficiently shared among the existing diverse model simulations and data processors [2, 3]. This situation is caused primarily by the lack of homogeneous, well-defined descriptions, transformations, and interconnections. Frequently, for model interoperability (or model coupling) a dataset generated by a given model (which runs as a software module) needs to be transformed into a different data format required by a subsequent module. The major issues in model coupling stem from the inconsistencies in the data formats and processing activities involved, combined with a lack of data conversion facilities for exchanging the results of running models. Moreover, during a sequence of model executions data analysis is usually needed in the form of advanced data transformations and processing, including the creation of 2D and 3D data visualizations.

Model coupling is a particularly complex topic with numerous solutions proposed, as for example described in [2], [3], and [23]. However, as explained in [9], most of the existing model interoperability methods require the users to modify source code in order to create meaningful model coupling scenarios. Furthermore, while there are several frameworks available that enable researchers to use model coupling techniques in a more systematic way only few solutions are accessible to a larger category of users and are fit for a wider variety of model simulations and data processing activities.

The work presented in this paper aims to address some of the limitations of current software tools used in climate change-related data processing and model simulation. The proposed solution aims to make steps towards increasing the software support for interconnected data processing and model simulation activities that pertain to climate change research. However, it is worth noting that the fundamental problems of model and data interconnectivity are similar in other scientific research fields. Hence, our proposed approach and supporting tools can be extended and applied to a variety of other fields’ data processes and model executions.

The challenges mentioned earlier demonstrate the need for modern software tools that support scientific research. Currently, various workflow-based tools for scientific work exist, but few deal with 3D environments, especially with immersive 3D environments. Furthermore, few allow a user-friendly, visual configuration of application scenarios (i.e., scenarios for running interconnected data processing activities and/or model simulations) with minimal or no code writing involved. Our work has focused on the development of a scenario management approach and associated software tools,
entitled the SUNPRISM framework, that integrates various kinds of data transformation (e.g., data format conversions) and model simulations using a flexible scenario configuration and execution process, automated code generation, and advanced capabilities for visual representations of scenario outputs [22].

The name SUNPRISM for our proposed solution consists of the combination of the words SUN and PRISM. The SUN part makes reference to the star at the center of the Solar System that makes all climates possible. The PRISM part suggests two analogies: first, similar to how an optical prism transforms the rays of light, the SUNPRISM framework is capable of processing and transforming sets of data; second, the prism is an eminently 3D geometrical shape, and one of the SUNPRISM framework’s distinguishing capability is that of displaying 3D visualizations, which inherently consist of 3D shapes.

Based on our previous work on CAVEMANDER [4, 17] and RTSIS [16], we propose the new and original SUNPRISM approach that allows the creation of application scenarios and their execution in 3D environments, including immersive virtual reality (VR) environments, by using a flexible and extensible workflow-based software development framework. This approach is supported by the SUNPRISM tools, which consist of two software applications developed for this project, the SUNPRISM Scenario Manager and the SUNPRISM Visualizer. The research contributions of the approach stem primarily from the combination it offers: (a) an end-user oriented process for visually designing and configuring various types of application scenarios that involves limited or no code writing; (b) a flexible workflow-based scenario construction and execution software environment developed using modern software engineering techniques; and (c) advanced visualizations of scenario outputs, including in immersive VR environments such as the CAVE.

This paper, in its remaining sections, is organized as follows: Section II presents an overview of the SUNPRISM approach, including its main phases and the specific modes of operation of its associated software tools, the Scenario Manager and the Visualizer; Section III provides details on the software tools’ architecture and user interface; Section IV illustrates using the SUNPRISM framework on an application scenario focused on retrieving forecast data from the National Digital Forecast Database (NDFD) web service and visualizing it using geographical 3D representations; Section V contains a short review of related work on scientific workflow tools, workflow-based visualization tools, and collaborative virtual environments; and Section VI provides the conclusions of the paper and outlines several directions of future work.

II. THE SUNPRISM APPROACH

The main goal of this work has been to provide a new methodological approach and related software framework, denoted SUNPRISM, capable of providing new, more effective and accessible ways of performing climate change-related simulations, including data processing, model execution, and model interconnection activities. SUNPRISM enables enhanced simulation activities, increased collaboration support for climate change education and research, and improved learning and training mechanisms. The first of the two major end products of the SUNPRISM approach is a visual, dataflow-based environment that allows user-friendly interactive creation and execution of simulation and data processing scenarios. A second major product associated to the SUNPRISM approach is a 3D data visualization tool with real-time modification and playback capabilities. The five modes of operation (or phases) characteristics for the SUNPRISM approach are as follows:

- **Object Registration**: Allows an end-user to register to the system new data resource objects and new computational activity objects, which may range from simple data processing activities to larger model simulation software applications. Some basic to more advanced programming is usually needed in this mode, depending on the complexity of the registered objects.

- **Scenario Configuration**: Provides the capability of visually defining data processing or simulation scenarios (in short application scenarios) by meaningfully interconnecting data and computational objects available in the system.

- **Code Generation**: Generates the associated executable code for a given application scenario. To achieve this, behind the scenes of the visual interface a software designed and implemented using modern DSL concepts has been developed.

- **Scenario Execution**: Runs the executable code associated with a given scenario, and based on user decision, stores the results obtained.

- **Data Examination**: Presents and allows visualization of data resulting from executing application scenarios. Notably, 3D immersive visualizations of data, including time-based presentations (animations) are possible.

An overview of the proposed approach and the interconnections between its modes of operation is presented in Figure 1, which depicts the high-level activity flow of the process outlined above. It is worth noting that this is only a high-level diagram describing the approach, and while the modes are further explained in this section, more details regarding the specifics of each mode of operation are provided in [22].

A. **Object Registration**

In this initial phase of operation various data resources and computational activities are created or uploaded and made available to the proposed SUNPRISM workflow environment through object registration. Both data resources and computational activities have software program modules associated with them, the difference between the two types of entities being that data resources are data-centered (e.g., consists of a dataset and its associated properties and queries) while computational activities are algorithm-intensive (i.e., they provide data processing or model execution capabilities). Object registration consists of specifying detailed properties, including the visual interface presentation, of the program module associated with either a SUNPRISM data resource
object or a SUNPRISM computational activity. Both types of objects have specific configurable parameters, where these parameters will be later evaluated to bind all specified information together to generate the customized executable code dynamically at run-time. Advanced constraints, such as potential input and output data resources for a computational activity object can be defined, which enhances the usability of the environment and facilitates the development of new program modules. The objects, either data or computations, can be added (registered) to the system, edited (updated), and removed from the SUNPRISM environment.

B. Scenario Configuration

The visual representation of SUNPRISM objects and their connections (using dedicated graphical symbols) is the distinguishing characteristic of the proposed approach and supporting workflow environment. During scenario configuration, via the environment’s user interface various application scenarios can be created by connecting registered objects (data resources and computational activities) based on the objects’ interface descriptions and other properties defined during their registration. An application scenario is in essence a data-driven execution workflow showing connected processes (computational activities) and the data exchanged between them. In a scenario, each object may have input connections from (and output connections to) more than one object, thus allowing the description of both sequential and parallel executions.

C. Code Generation

In this phase of the SUNPRISM approach, executable code for a given scenario is generated automatically using software engineering techniques related to the DSL methodology. Specifications of data resource objects, computational activity objects, and scenarios are defined by their abstract syntaxes, or meta-models, whereas a scenario is constructed accordingly (based on its constituents objects syntaxes) in the earlier scenario configuration step. During the code generation phase, an application scenario, which can be considered an instance of the meta-model for a scenario type, is parsed by the SUNPRISM code generator, which recognizes the scenario’s semantics based on its meta-model definition and transforms the scenario’s description into lower-level executable code.

D. Scenario Execution

In the scenario execution phase the generated executable code associated with a scenario can be run in the proposed SUNPRISM workflow environment by running its constituent processes and data transformations in the sequence specified by the scenario’s description and according to the operational (run-time execution) characteristics of all the data resources and computational activity objects involved in the scenario. While running a scenario the users are able to view the progress of its execution and obtain information on the current state of the scenario’s objects.

E. Data Examination

A scenario execution obtains input data from specified data resources and transforms it into output data based on the specification of program modules associated to computational activity objects. Output data must also be one of the data resource objects registered and recognized in the SUNPRISM software environment. In the data examination phase of the SUNPRISM approach the users are mostly interested in viewing and analyzing the output data that has resulted from running the scenario execution that they have configured. Thus, the data examination phase of the proposed approach is a critical operation in the SUNPRISM framework. While the examination of some data types may be achieved using external specialized tools that can be connected with SUNPRISM, such as spreadsheet programs, the new SUNPRISM framework also provides advanced capabilities for 3D data visualization, including dynamic data presentation and step-by-step playback. Notably, data visualizations in SUNPRISM can be performed in immersive environments such as CAVE. These facilities enable sharing scenarios and collaboration among scientists on visualizing and interpreting their results.
As stated previously, the SUNPRISM framework consists of two tools, namely the SUNPRISM Scenario Manager and the SUNPRISM Visualizer. The first component, the Scenario Manager, has been developed from scratch to provide a user-friendly visual interface for object registration, scenario configuration, code generation, and scenario execution. The Scenario Manager tool allows end users to create and execute workflows using the dataflow paradigm. In addition, the Scenario Manager allows researchers or developers to add new data resources and computational activities to the framework. The second component tool, the Visualizer, has been developed to support the examination of various types of output data, including through immersive interactive presentations based on 3D scenes. In essence, the SUNPRISM framework incorporates the Scenario Manager to enable workflow executions that transform data by running computational activities, and integrates the Visualizer to allow the visualization of complex data outputs in 3D. Figure 2 presents the overview of the SUNPRISM framework.

It is worth noting that both the Scenario Manager and the Visualizer are standalone applications that can be operated independently. The main purpose of integrating these two tools in a single framework is to bridge the data processing and model execution activities supported by the first tool with complex 3D data visualizations enabled by the second. To this end, we have created a tool loader for the integrated SUNPRISM framework tools, shown in Figure 3.

A. SUNPRISM Scenario Manager

The main window of the SUNPRISM Scenario Manager, shown in Figure 4, consists of five components; the menu bar and toolbar at the top of the window provide accesses to all the available operations, object panels on the left (dedicated to data resource objects and computational activity objects) allow insertion of objects in the scenario drawing area (the central panel), and the property table on the right allows inspection and configuration of each object included in a given scenario.

B. SUNPRISM Visualizer

The main window of the SUNPRISM Visualizer, shown in Figure 5, consists of four components; similar to the SUNPRISM Scenario Manager, the menu bar and toolbar at the top of the window provide accesses to all the available operations, the scene objects panel on the left allows the insertion of scene objects in the scene drawing area (the central panel), and the property table on the right allows the inspection and modification of each object included in a scene.
C. Scenario Execution

A scenario can be constructed in the Scenario Manager with objects and connections, and once the scenario is configured with parameters it can be executed within the tool. The execution order is determined based on the scenario’s dependency relationships (interconnections between data resource objects and computational activity objects). In other words, all object instances that provide incoming data to an object instance must be executed prior to running this instance. A scenario can thus be considered a directed acyclic graph with jobs as nodes and dependencies as edges, which can be sorted topologically using a depth-first search as described in [6], assuring the execution with a correct order. Once the execution order is defined, the executable code is generated automatically to bind associated scripts of objects altogether. The Scenario Manager executes the generated code sequentially, and produces the expected output data, which can be then examined after completing a scenario’s full execution. Figure 6 shows a schematic depiction of SUNPRISM’s automatic code generation and output production. Note that in this figure, the central resource, an instance of gsod_csv_to_prismvis, is a computational activity, all other being data resources.

IV. Application Scenario

An application scenario was developed to retrieve the forecast data from National Digital Forecast Database (NDFD) through their provided web service. The NDFD consists of forecasts of weather elements from nationwide high resolution measuring points, in which the data can be accessed by various methods, including HTTP, FTP, and web services, provided by NWS Digital Services [20]. Moreover, the application scenario was developed to produce 3D visualization outputs that can be used to display retrieved and processed data in a three dimensional view, where the scene is created using the world map and vertical bars are indicating specific climate information at corresponding geographical locations.

The scenario for this application contains a data resource object (web_service_in) that retrieves measurements data from the web service as input, which connects to a computational activity object (conv_ndfd_to_prismvis) that transforms this input data to a format that can be readily loaded to an interactive 3D visualization environment (prismvis). The same output goes also to an image data resource (image_out). The configured scenario for this application example is shown in Figure 7.

Once the scenario configuration and execution is completed, the output files are stored at the locations specified by the objects’ parameters, where they are available and ready for data examination. The output files can be opened in tools associated to them for further inspection. In this application scenario, and to demonstrate one of the innovative features of the SUNPRISM framework, a prismvis 3D visualization is associated to, and can be opened in the SUNPRISM Visualizer, as shown in Figure 5 (included earlier in the paper). In this example, the world map is rendered at a ground level, and a vertical bar is placed at the geographical location of each point returned by the NDFD web service query, where the height of each bar represents a specific data, for example the maximum temperature forecast at the time specified in the scenario. In addition, each object in a scene, a bar in this visualization, can be selected to view associated properties and values for the object, thus facilitating interpretation and analysis of data (Figure 8).

V. Related Work

The distinguishing characteristics of the proposed SUNPRISM Framework include providing software tools that allow the construction and execution of scenarios as computational workflows, as well as the examination of complex scenario output data through interactive 3D visualizations. Furthermore, in SUNPRISM such visualizations can be projected in VR environments such as the CAVE, in which users can interact with the same scene from different perspectives, thus allowing communication and collaboration in virtual shared environments. These are unique features that
make SUNPRISM stand apart in the landscape of tools for model and data interoperability. Due to space limitations, the following is only a succinct review of several tools and approaches similar to SUNPRISM, grouped in three categories: scientific workflows, workflow-based visualizations, and collaborative virtual environments. More details of related work, including a criteria-based comparison are available in [22].

A. Scientific Workflow Environments

Kepler is an open source scientific workflow software environment which allows users to design and configure workflows that manipulate and analyze scientific data [14, 15]. It provides visual representations of workflow components and their connections, and an interactive interface to move objects around and to specify linkings between objects. Kepler is a powerful and complex visual programming environment that allows execution of a broad range of workflow-based scenarios. Another well-known open source workflow management system is Taverna, which emphasizes a data flow centric model that integrates and analyzes heterogeneous and semi-structured data from various service-based resources [24, 28]. Several other workflow management systems have been actively developed, each of them with emphasis on and strengths in different areas. For example, Discovery Net [25] and Triana [29] are among the most common titles in the workflow communities. A comparison of notable workflow systems, including Discovery Net, Triana, Kepler, and Taverna, can be found in [8].

B. Workflow-based Visualization Environments

In the category of workflow-based visualization tools, SCIRun is a visual programming environment that can be used to construct and interact with large-scale scientific computational simulations [21, 26]. Available as open source, it is a “computational workbench,” where scientists can design and run dataflow-based algorithmic simulations, and “steer” the components’ configurations while they are still in progress. Another open source visualization pipeline system is VisTrails, with advanced provenance management features [27, 30]. The VisTrails system provides a visual dataflow programming interface and allows the construction of workflows with a scripting interface. Like SCIRun, modules are networked together as a dataflow, which triggers the execution of the computations with data passings. In addition, its advanced provenance management capabilities facilitate scientific explorations and collaborations. However, VisTrails is heavily focused on graphics, rather than on items for customized simulations such as data resources. Additional information on workflow-based visualization environments is available for example in [13].

C. Collaborative Virtual Environments

Numerous libraries and frameworks have been developed to create immersive VR applications, with the CAVE being probably the most immersive virtual environment [7]. CAVELib is a sophisticated API library that allows developments of VR CAVE applications [6]. It provides a bridge between various hardware devices, allowing developers to access advanced rendering and communication features. Since its creation, it has inspired the development of several similar APIs. For example, FreeVR is a CAVE application development library that provides features and capabilities similar to CAVELib’s, facilitating the integration of various input and output hardware devices [10]. Hydra, inspired by FreeVR, is another API library that facilitates the development of interactive virtual reality applications, allowing various types of input devices and output displays to be integrated seamlessly [11]. Hydra is designed to be a cluster-based interface library for virtual reality applications with multiple systems over the network, which improves the performance with the cluster-based memory management and the efficient communication between input and rendering modules [11]. VRJuggler is also one of the widely used toolkits, which provides a virtual platform, a cross-platform environment to design and execute virtual reality applications independent to hardware configurations and operating systems [1, 31]. VRJuggler emphasizes an object-oriented approach and device abstractions to ease the application development process and
execution. Our own CAVEMANDER offers a framework that incorporates a software engineering method in CAVE-based VR application developments [4, 5, 17]. More recently, RTSIS has extended the CAVEMANDER framework by focusing on the real-time supervisor intervention paradigm as applied in CAVE-based applications [16]. RTSIS was developed as a desktop tool that allows creating scenarios visually, controlling scenario executions, and monitoring and intervening (by a supervisor) in real-time upon the actions of a human user (typically, a trainee) acting in CAVE. RTSIS significantly extends the CAVEMANDER framework by adding another software subsystem with capabilities that facilitate rapid scenario creation, scenario modification, and scenario execution, and also by introducing the supervisor’s intervention paradigm. RTSIS represents our prior (yet recent) work on which SUNPRISM has been built.

VI. FUTURE WORK AND CONCLUSION

In this paper we have presented the new SUNPRISM approach and associated software tools for application scenario creation and execution in support of climate change research and experimentation. Such application scenarios include a combination of data resource and computational activity objects that show the capabilities of SUNPRISM for supporting scientific research work in climate change-related disciplines, as well as in other disciplines. Nevertheless, there are still many directions of future work that could be beneficial to SUNPRISM users. Such directions include:

- Additional functions and capabilities for the SUNPRISM Scenario Manager, for example a help information system with details on how data resource and computational activity objects can be used. Also useful would be a history trace of user interactions, including details of scenario modifications and outputs from scenario executions. Furthermore, there is also work needed to facilitate the sharing of the registered objects (e.g., via an online repository), so that researchers can more easily disseminate their modeling and simulation work.

- Additional functions and capabilities for the SUNPRISM Visualizer, for example the ability to integrate other types of external display systems, as it currently supports only Hydra-based CAVE displays for external projections. In addition, the communication model of the Visualizer with other display systems has room for improvement, for example for integrating multiple projection screens.

- Thorough testing and open source release, with involvement of various testers and end users. The feedback obtained will be incorporated in the development of a revised version of the SUNPRISM Framework.

- SUNPRISM adaptation to other research domains, for example to biomedical engineering or physics.

The main contribution of the work presented in this paper (and detailed in [22]) is the proposal of the new SUNPRISM approach and supporting tools for data transformation and model simulation, with applications to environmental science research. The distinguishing characteristics of SUNPRISM are:

- A user-oriented process for visually designing various types of application scenarios with limited or no coding;
- A flexible workflow-based scenario construction and execution software environment developed using modern software engineering techniques; and
- 3D visualizations of scenario outputs, including in immersive VR environments such as CAVE.

The combination of the above characteristics makes this approach unique and it is where its novelty primarily resides. Furthermore, SUNPRISM fosters research and development collaboration, as it offers the appropriate instruments (method and tools) for scientists and other interested users to share their research and experimentation scenarios and examine together the results of these scenarios.

An application scenario focused on retrieving and visualizing data from the NDFD web service has also been presented to illustrate the approach in practical settings. While the SUNPRISM Framework is currently used for running scenarios pertaining to model coupling in climate research it can be utilized without modifications in several other application domains, for example in command and control or other 3D-based training and simulation applications. For a number of additional domains only minimal and straightforward adaptations will likely be needed.

ACKNOWLEDGMENT

This work was made possible through the support provided by the National Science Foundation under Cooperative Agreement No. EPS-0814372.

REFERENCES


[10] FreeVR. Available at: http://www.freevr.org
    [Accessed November 18, 2011]

    [Accessed November 18, 2011]

[12] Idaho EPSCoR. Available at: http://www.webs.uidaho.edu/epscor
    [Accessed November 1, 2011]


[14] Kepler. Available at: https://kepler-project.org
    [Accessed November 14, 2011]


    [Accessed November 1, 2011]

[19] New Mexico EPSCoR. Available at: http://www.nmepscor.org
    [Accessed November 1, 2011]

    [Accessed November 8, 2011]


    [Accessed November 14, 2011]


    [Accessed November 16, 2011]

[29] Triana. Available at: http://www.trianacode.org
    [Accessed November 16, 2011]

    [Accessed November 14, 2011]

[31] VR Juggler. Available at: http://vrjuggler.org
    [Accessed November 18, 2011]