# DRIHM2US

## DISTRIBUTED RESEARCH INFRASTRUCTURE FOR HYDRO-METEOROLOGY TO UNITED STATES OF AMERICA

**Minute of the workshop B1: "Using Research Infrastructures for Integrated Environmental Modeling"**
**at the 7th International Congress on Environmental Modelling and Software (iEMSs)**
**June 15-19, 2014, San Diego, California, USA**
**(http://www.iemss.org/sites/iemss2014/)**

*DRIHM2US (G.A. n° 313122) is co-Funded by the EC under 7th Framework Programme*

# Minute of the iEMSs Workshop B1: Using Research Infrastructures for Integrated Environmental Modeling

Place: San Diego, California, US

Duration: Tues 17th June, iEMSs 2014, San Diego

Editors: Quillon Harpham

Agenda:

9:00 – 9:05 Welcome and Introduction
9:05 – 9:20 Hydroshare Demonstration
9:20 – 9:35 FluidEarth Demonstration
9:35 – 9:40 DRIHM MAP
9:40 – 10:20 Discussion

Attendees:

Quillon Harpham (HR Wallingford), Bert Jagers (Deltares), Elisabetta Fiori (CIMA Foundation), Antonella Galizia (CNR-IMATI), Vladimir Dimitrijevic (RHMSS), Ljiljana Dekić (RHMSS), Leonard Levin (EPRI), Peter Gijsbers (Deltares), Becky Schmidt (CSIRO Land and Water), Vaclav Petras (NCSU), James Syvitski (CSDMS), Mark Bryden (Ames Lab), Holger Dettki (VC-WRAM), Lucy Bastin (JRC), Rick Hooper (CUAHSI), Robert Argent (BoM), James Droppo (PNNL), Gene Whelan (US EPA), Sandy Elliot (NIWA), Carlo Giupponi (UNIVE), Fedor Baart (Deltares), Beaufils Xavier (LTHE), Shaun Livingston (BYU), Veronique Chaffard (LTHE), Jeff Sadler (Brigham Young University), David Tarboton (Utah State University), Tom Purucker (US EPA).

Apologies: Ilya Zaslavsky (SDSC).

The workshop began with demonstrations of HydroShare and FluidEarth (OpenMI). A set of seed questions was posed, structured around the DRIHM MAP concept (see presentation available on the DRIHM2US site). This was intended to set context and give a summary of the latest position on research infrastructures for integrated environmental modelling. There were no dissenting voices to this summary, indeed, rather than directly respond to the questions posed, the discussion immediately proceeded along a more philosophical line with the group considering the main issues arising from the current positions of such infrastructures and how to characterise and overcome the challenges going forwards. Rarely does a group get together with such broad and deep collective experience in this subject and the discussion had considerable richness and insight. Accordingly, it is represented here in an attempted commentary.

The core theme of the discussion was in the division of responsibilities along the entire supply chain – from writing core model engines, to creating instances, to integrating with other models, to running and using results. It seems that optimal success will be derived by attributing these responsibilities successfully. More than 10 years' work has gone into getting where we are today and the trends are becoming clearer.

**The Technical Burden**

We are trying to create an environment where scientists can do 'science' without being hampered by 'computer programming' issues. There is a convergence of attention on interfaces so that between these interfaces there is freedom to operate. Standards such as OpenMI and BMI have come to similar conclusions in balancing the burden between model developers and integrators. One such enabler of this is to allow the equation solvers themselves to be unhampered by the management code associated with integration or model running. Moreover, 'Software as a Service' (SaaS) removes platform issues from the modellers completely. However, there is a potential tension between standards and freedom: standards must enable successful creativity and not hamper it.

As these technologies find their way closer to high volume commercial use, practitioners are finding that their customers are not interested in the technologies used to create a better answer (even if they are open source), they simply want to buy a better answer. The weather forecast is an excellent example of such a service where the improvements are largely hidden from the customers.

**User Interfaces**

Do we need standards for user interfaces? Do all users benefit from user interfaces since they can restrict active, iterative development of the core model code base? Models such as ROMS and TELEMAC do not necessarily require user interfaces since expert users prefer to have ongoing access to the code in order to address a continual desire for new use cases. However, many typical use cases can be addressed more economically by use of graphical user interfaces (GUIs), indeed any sort of re-use of models by practitioners other than the developer emphasises the need for GUIs. This rightly places the burden of understanding the base model and implementing the GUI on the developer so that others need just learn the user interface. Certain activities such as formulating input datasets, creating boundary conditions or calibrating models do not lend themselves to use of GUIs (which would be more typically used in mature models). The tasks are too complicated and too much restriction is applied. Indeed, a GUI is in itself a standard and the same trust and control issues arise as with other standards.

Ultimately, model operation through the command line gives optimal control. Use of the command line is necessary for batch running of models through infrastructures, especially when ensembles are required and there is clearly a necessity for both command line and GUI operation. If implemented correctly with a component software architecture there is no conflict between them. Moreover, dedicated GUIs for suites of models can develop into unified integrated modelling platforms.

**Future Paradigms**

So what will the next dominant paradigm be? We have progressed from the slide rule to the pocket calculator to modern computers. We have decades of experience of numerical modelling:

- Practitioners tend to agree that a new role is forming in terms of integrated modelling – that of 'model integrators' or 'curators' whose expertise leads to evaluating valid combinations of models and the issues which will arise in the use of the combination.
- When models are used as commodity tools (knowledge encapsulators) the end user may simply require 'the answer', but when models are used for hypothesis testing the end user is the modeller testing the hypothesis. Distinguishing this allows model

frameworks to be tailored appropriately for these (and any other) valid uses.

- Controlled vocabularies are necessary and desired. The assumption that they will all use the English language may be disenfranchising non-English speakers.
- There is also potential in building intelligence into integrated modelling systems, leading the user into the best courses of action for their use case. For example, different competing models will have different strengths and weaknesses: better results at the equator / better results at the poles; a high sensitivity to incomplete supporting data / a low sensitivity to incomplete supporting data.
- As standards tend to operate at component interfaces, it is desirable that implementing any standard should minimise its intrusiveness to each component. This can be solved at two levels, the library level and the object level. If this lack of intrusiveness is achieved then technology can be exported across technical domains such as inserting the computational core of a numerical model into a game engine.

**Notes from Hydroshare Demonstration**

HydroShare: A web-based infrastructure to share hydrological data and models – discover, access, retrieve, analyze

- Digital divide: research flow chart – "archane" HPC details
- Everything is a resource: data discovery tool, data loader, analysis/visualization tools, model à one repository
- Current status: generic file sharing functionality, more custom "hydrology" specific types to be added
- http://beta.hydroshare.org
- Moved from Drupal to Django based
- Permanent links -> submit for publication should provide DOI: moving to CUAHSI for "perpetuity"
- No checks for generic resources, but for time series WaterML x (2.1?) will be required
- Social interaction elements

**Notes from FluidEarth Demonstrations**

Going through one training example from http://elearning.fluidearth.net

- Flow from western pond to eastern pond
- East boundary of western pond (10x10 grid cells) aligned with western boundary of eastern pond (5x5 grid cells); spatial mapping at boundary
- Water level slope in left and right ponds
- Do the models need to have the same time steps? Pipistrelle includes adapters for different time steps, but researcher needs to determine whether it's useful.
- No smart features in these simple demo models