International Conference on Hydroinformatics

2014

DRIHM - An Infrastructure To Advance Hydro-Meteorological Research

Bert Jagers

Antonio Parodi

Michael Schiffers

Nils Otto vor dem Gentschen Felde

Christian Straube

See next page for additional authors

Authors
Bert Jagers, Antonio Parodi, Michael Schiffers, Nils Otto vor dem Gentschen Felde, Christian Straube, Andrea Clematis, Daniele D'Agostino, and Quillon Harpham

DRIHM - AN INFRASTRUCTURE TO ADVANCE HYDRO-METEOROLOGICAL RESEARCH

BERT JAGERS (1), ANTONIO PARODI (2), MICHAEL SCHIFFERS (3), NILS OTTO VOR DEM GENTSCHEN FELDE (3), CHRISTIAN STRAUBE (3), ANDREA CLEMATIS (4), DANIELE D'AGOSTINO (4), QUILLON HARPHAM (5)

(1): Deltares, P.O. Box 177, 2600 MH, Delft, The Netherlands

(2): CIMA Research Foundation, University Campus of Savona, Via Armando Magliotto 2, 17100 Savona, Italy

(3): Ludwig-Maximilians-Universität München, Munich Network Management Team (MNM), Institut für Informatik, Oettingenstr. 67, D-80538 München, Germany

(4) Inst. of Applied Mathematics and Information Technology – Consiglio Nazionale delle Ricerche, Via De Marini 6, 16149, Genoa, Italy

(5) HR Wallingford, Howbery Park, Wallingford, Oxfordshire, OX10 8BA, United Kingdom

One of the main challenges in hydro-meteorological research (HMR) is predicting the impact of weather and climate changes on the environment, society and economy, including local severe hazards such as floods and landslides. At the heart of this challenge lies the ability to have easy access to hydro-meteorological data and models, and facilitate the collaboration across discipline boundaries. Within the DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology, http://www.drihm.eu, EC funded FP7 project 2011-2015) we develop a prototype e-Science environment to facilitate this collaboration and provide end-to-end HMR services (models, datasets, and post-processing tools) at the European level, with the ability to expand to global scale. The objectives of DRIHM are to lead the definition of a common long-term strategy, to foster the development of new HMR models, workflows and observational archives for the study of severe hydro-meteorological events, to promote the execution and analysis of high-end simulations, and to support the dissemination of predictive models as decision analysis tools. For this we implement a service portal to construct heterogeneous simulation workflows that can include deterministic and ensemble runs on a heterogeneous infrastructure consisting of HPC, grid and Windows cloud resources.

INTRODUCTION

Floods may occur as result of rising water levels along the coast due to storms and hurricanes, local intense rainfall, or from high river discharges due to upstream accumulated rainfall or snowmelt. In particular intense precipitation in mountainous catchments can lead to a rapid increase in river discharge causing flash flooding with large damages and deadly consequences. Table 1 lists a number of such intense events that happened around the globe over the past few years. Predicting such events is difficult: in a few hours the discharge may increase rapidly to peak value, giving little lead time when flood predictions are to be based on observed rainfall rates and measured discharges. This is in contrast to for instance large discharges due to accumulated snowmelt throughout the Rhine River catchment, which may be estimated well ahead of time. The only option to increase the lead time for flash flood predictions is to include a weather model in the forecasting chain. This, however, introduces additional sources of uncertainty: are we predicting the right amount of precipitation, and are we predicting it on the correct side of the mountain (affecting the catchment that it will flood). This is a field in which still a lot can be learnt and improved.

Table 1. A selection of recent flash flood events

Date	Location	Nr of casualties	Reference
4 Nov 2011	Genoa, IT	6	[6]
20 Jun 2013	Alberta, CA	4	[5]
9 Sep 2013	Colorado, US	8	[1]
18 Nov 2013	Sardinia, IT	18	[2][3]

Within the EC funded FP7 DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology[1], 2011-2015) we develop a prototype e-Science environment to facilitate the necessary collaboration across hydro-meteorological discipline boundaries. The DRIHM platform allows users to run simulation workflows consisting of a number of meteorological, hydrological, hydraulic and impact models across a Europe-wide heterogeneous computing infrastructure that includes HPC, grid, and Windows cloud resources.

Integrating such a heterogeneous group of models and computational resources combined with challenges of data transfer between them due to the distributed nature requires an iterative approach especially since availability and type of resources may be shifting over time. We followed therefore a pragmatic “learning by doing” approach. The following sections give an overview of the general architecture of the platform, the standards and models currently implemented, and a brief description of the portal and services provided so far.

DRIHM ARCHITECTURE

Following an initial analysis of the workflows currently used by the project partners in house to run individual or limited chains of models, we have developed a prioritized list of requirements subdivided in the categories: data management (such as storing, converting and serving gridded and point series data), model discovery and setup services (such as identifying and setting up domain specific models), model management (such as running individual models on different types of hardware), workflow management (such as building workflows and connecting model outputs and inputs), and post processing (such as presenting and exporting simulation data). To prioritize the developments, we work in the same direction “as the water flows” through a set of three experiments: the first one has a focus on meteorology, via the second one which links meteorology to hydrology, to the final experiment that links all components together with a focus on obtaining hydraulic and impact predictions.

Linking data, models and tools is challenging by itself – the more so if the models vary from HPC to Windows based. From different perspectives various groups around the world

[1] http://www.drihm.eu

have developed coupling approaches [4], but none of them has been tried and tested for the whole range of models and computational resources relevant for our DRIHM platform. Luckily the processes investigated don't need all to be coupled in two directions. In flash flood situations, the meteorological situation leads to precipitations, which after saturating the soil leads to runoff, it's only at the point that the flood waters start to flood from the river into the floodplains and the cities that two-way interaction is crucial. Therefore, we have chosen to work in DRIHM with file-based one-way coupling between the meteorological and hydrological models, as well as between the hydrological and hydraulic models, while the hydraulic (and impact) models may run in a two-way coupled composition.

Figure 1 contains a sketch of the current architecture with examples of services used and provided (the set of post-processing options on the right hand side is still largely under development at the time of writing). The user interacts with the system via the portal, or science gateway, on the left (http://portal.drihm.eu). It allows you to select and configure a chain of models; model configuration (also referred to as model set-up) may occur by specifying the model parameters directly within the portal, or by preparing model input files offline and uploading them via the portal. During the configuration of a workflow composed of multiple models, the portal may perform a number of simple checks to verify the consistency of the selected model configurations, such as time span, time steps and spatial domain. These checks will be performed based on metadata of the simulation engine, the model (base schematization) and the configuration files created or provided.

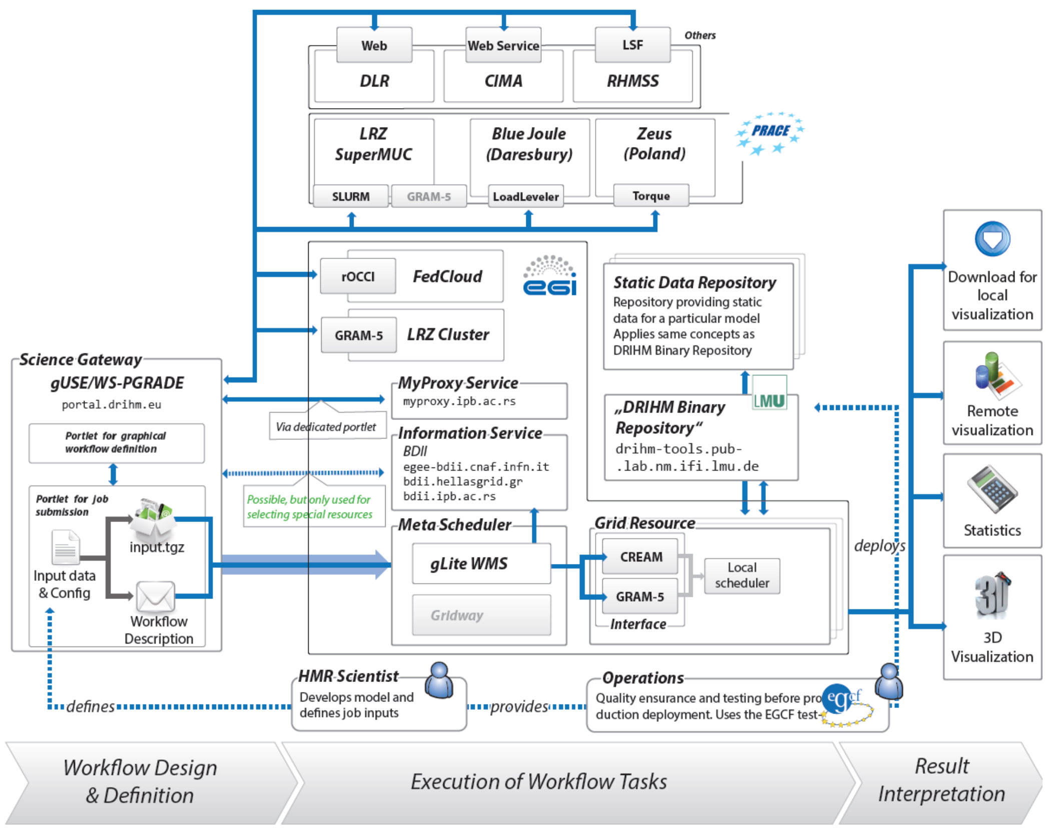

Figure 1. Overview of the current DRIHM architecture

DRIHM WORKFLOWS

Once the models have been configured and a workflow has been assembled, the workflow can be submitted for processing by the infrastructure. Each model component may be run on one or more of the computing systems shown. For high-resolution weather or hydraulic models the supporting hardware may be an HPC resource, such as a PRACE or local supercomputer or a high-performance grid resource; for other models that may be a different resource: the portal has been configured to know which component runs best on which system. This configuration step is one element in the DRIHMification process.

The DRIHMification process is inspired by the gridification process necessary for running software on “the grid”. In order to run a program on the grid, it must be made as independent of the hardware and software configuration as possible. This means that all software components necessary for running a model (except for the base system libraries) have to be provided together with the main program. This usually includes a series of compiler specific shared libraries, but may occasionally also include runtime environments for scripting languages such as Python, or a JAVA (or even Windows) virtual machine. After removing any hardcoded paths such that the program can run in any folder configuration, the program needs to be wrapped with shell scripts to support stage-in and stage-out processing of the input and output data respectively. To run a program within DRIHM we require in addition that the software is uploaded to the “DRIHM Binary Repository”, that generic metadata on the software is provided (for documentation harvesting and processing by the portal), and that the program directly (or indirectly via a wrapper) satisfies a number of interface standards imposed by DRIHM. The models currently available on the DRIHM infrastructure are: WRF, RainFARM, DRiFt, and RIBS, while work is ongoing for WRF-NMM, Meso-NH, Mascaret, Delft3D, an OpenStreams-PCRaster HBV implementation, and a number of other models. We aim to have a good variety of open source models available on the DRIHM platform, but we don’t exclude closed source commercial models. However, such models add to the complexity and sustainability risks of the platform: it is unlikely that in all cases continued funding for the necessary licenses will be found, such that the more likely scenario becomes that users should be able to use their own license for those models on whatever system they will execute.

Portal

Since the models in a single workflow may need to be run on different hardware platforms, the portal must be able to interact with all these platforms. Since the protocols are diverse, we build on the generic gateway gUSE framework with the Liferay-based WS-GRADE frontend developed by the EC FP7 SCI-BUS[2] project. In collaboration with those developers the system is extended and tuned for DRIHM. Furthermore, the portal has to effectively deal with the variation in access policies and user credential management by the different hardware centers. This is will remain a challenge for both portal developers and end users.

Support information and services provided by others, such as the Cb-TRAM service for identification and tracking of thunderstorms based on satellite data hosted by the DRIHM partner DLR (German Aerospace Center), are seamlessly embedded in the portal for a consistent look and feel. The Cb-TRAM service allows users to analyze the development of storm clouds for a number of pre-selected events – including an option to download derived storm cloud contours into Google Earth.

[2] http://www.sci-bus.eu

Binary Repository

To simplify distributing and updating models throughout the heterogeneous infrastructure, a central binary repository was created that hosts all binaries to be run throughout the grid infrastructure. The repository is split in a git maintained core repository for testing, development and branching, and a production area that is publicly available as read-only. The files in the production area are automatically updated when new models are released for public use by the development and maintenance teams. The first step in running a model is downloading or updating a local mirror of the production area via rsync command; this step ensures that always the latest version of the software is used.

Standards

In the DRIHM infrastructure, we currently impose three standards for data exchange. Firstly, the P-interface focuses on the exchange of data from meteorological to hydrological (and in the future possibly oceanographic) models. The main element in the P-interface is currently the precipitation, hence the name, but for some models downstream in the workflow other quantities are relevant as well. For this exchange we require a netCDF v3 file using CF v1.6 conventions in line with the OGC netCDF-CF standard; the file should contain the quantities listed in Table 2 on the next page to be compliant with the P-interface.

The second standard concerns the exchange of data from hydrological to hydraulic models and is referred to as the Q-interface. As the name suggests, the main (and currently only) element in the Q-interface is the discharge. For this exchange we require a WaterML 1.1 file in line with the OGC standard. Every hydrological model implemented on the DRIHM platform should be able to use the precipitation data from the P-interface netCDF file as obtained from the meteorological model, and should be able to produce discharge time series at one or more points in WaterML format for the Q-interface.

Finally, we are using OpenMI for the exchange of data amongst the hydraulic and impact models. For this the OpenMI 2.0 interface is used, which has been submitted to OGC for adoption. It should be noted that these exchange standards address only the data needed that needs to be transferred between models; users may want to inspect much more data and hence all models are allowed to write out results in whatever file format they like. Such additional output will be treated as a black box by the DRIHM infrastructure.

Table 2. Quantities required for the DRIHM P-interface netCDF file.

Quantity (CF standard_name)	Unit	Height (above surface)
lwe_thickness_of_precipitation_amount	m	(0m)
lwe_thickness_of_stratiform_precipitation_amount	m	(0m)
lwe_thickness_of_convective_precipitation_amount	m	(0m)
air_temperature	K	2m
specific_humidity	1	2m
surface_net_downward_longwave_flux	W m-2	(0m)
eastward_wind	m s-1	10m
northward_wind	m s-1	10m
surface_air_pressure	Pa	(0m)

OUTLOOK

The DRIHM project was started from a vision to develop a distributed research infrastructure that would give more computational freedom to individual researchers by breaking down the barriers of high performance hardware access and solving technological barriers for interoperability. The journey has been challenging as the software in the HMR research community has proven to be more diverse than the ICT partners had originally anticipated, and the ICT infrastructure is less uniform and more diversely standardized than the HMR researchers had expected. This means that more challenges had to be solved, and obviously still many challenges remain. Within the DRIHM project we continue until the beginning of 2015 to develop the prototype e-Science environment that we set out to implement. In the meantime, we are broadening our scope. Via the FP7 project called DRIHM2US[3] we collaborate with the NSF funded SCIHM[4] project to build a wider international collaborative network and perform tests with workflows that cross the Atlantic. And, last but not least, we are looking for ways to expand the user base for the DRIHM environment by extending into neighboring domains such as extending from weather to climate, and from river to coast.

REFERENCES

[1] Pearce, M. (September 23, 2013). "Number of missing after Colorado flood dwindles to 6". *Los Angeles Times*. Retrieved from http://www.latimes.com/nation/nationnow/la-na-nn-colorado-floods-missing-20130923,0,7588421.story on April 2, 2014.

[2] Price, M. (November 19, 2013). "Sardinia hit by deadly Cyclone Cleopatra", *BBC News Europe*. Retrieved from http://www.bbc.com/news/world-europe-24996292 on April 2, 2014.

[3] Masters, J. (November 20, 2013). "Floods from Mediterranean Storm 'Ruven' Kill 18 in Sardinia", *Weather Underground*. Retrieved from http://www.wunderground.com/blog/JeffMasters/floods-from-mediterranean-storm-ruven-kill-18-in-sardinia-italy on April 2, 2014.

[4] Jagers, H.R.A. "Linking Data, Models and Tools: An Overview", *iEMSs 2010 International Congress on Environmental Modelling and Software – Modelling for Environment's Sake, Fifth Biennial Meeting*, Ottawa, Canada, (2010).

[5] CTV News (June 26, 2013). "Residents of flood-ravaged Alberta community wait to return home". Retrieved from http://www.ctvnews.ca/canada/residents-of-flood-ravaged-alberta-community-wait-to-return-home-1.1342233 on April 6, 2014.

[6] Pisa, N. (November 5, 2011). "Six killed in floods as muddy torrent sweeps through the streets of an Italian city", *Daily Mail Online*. Retrieved from http://www.dailymail.co.uk/news/article-2057917/Italy-floods-Six-killed-muddy-torrent-sweeps-streets-Genoa.html on April 6, 2014.

[7] Galizia, A., et al. "Towards an interoperable and distributed e-Infrastructure for Hydro-Meteorology: the DRIHM project", *iEMSs 2014 International Congress on Environmental Modelling and Software – Bold Visions for Environmental Modelling, Seventh Biennial Meeting*, San Diego, USA, (2014).

[3] http://www.drihm2us.eu

[4] http://www.scihm.org