DRIHM2US

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

Final Report

Abstract: This document presents a summary of the analysis and guidelines resulting from the activities carried out in the project.

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

Document Information Page

Contract Number	313122
Project Name	Distributed Research Infrastructure for Hydro-Meteorology to United States of America
Project Acronym	DRIHM2US
Deliverable Number	-
Deliverable Name	Final Report
Work Package Number	4
Work Package Name	Dissemination
Deadline	31/03/2015
Version	1.0
Dissemination Level	PU
Nature	R
Lead Beneficiary	IMATI

Document History

Date	Version	Description
1st Mar 2015	0.1	Initial write up for review by the project team.
31st Mar 2015	1.0	Final version

Table of Contents

1 Executive Summary

The DRIHM2US project aims at advancing the scientific collaboration on both sides of the Atlantic towards improving the predictive ability of severe storms and utilization of these predictions for hazard prediction and control under climate change effects.

Such ambitious goal is based on a better understanding of the utilization of e-Science Infrastructures (e-Infrastructures) for these aims. The present report, although not an official deliverable, summarize the analysis performed during the project and the resulting guidelines.

2 Introduction

The DRIHM2US project has brought ICT and HMR research groups from Europe and US together on improving the interoperability, accessibility, and usability of e-Science infrastructures (e-Infrastructures) for the Earth Sciences research with a primary focus on High Impact Weather Events (HIWE) such as the recent floods in Genoa, Italy (2011, 2014) and Boulder, Colorado, USA (2013).

The project was able to determine that the issues being faced by the USA and Europe are similar (or at least closely related) as, indeed, are the solutions being attempted. To improve the prediction skill for such extreme conditions there is a general agreement on the need to increase the resolution of meteorological simulations, support multi-model ensembles, construct most probable precipitation scenarios by comparing simulation results with latest remote sensing and in-situ data sources, and run ensembles of hydrological, hydraulic, and impact models with appropriate levels of detail. In short, there is the need to make effective use of modern ICT infrastructure to combine all data sources into the best forecasts as this will help to manage such situations more effectively and save lives.

The DRIHM (Distributed Research Infrastructure for Hydro-Meteorology, www.drihm.eu) project in Europe and the SCIHM (Standards-based Cyberinfrastructure for Hydro-Meteorology) project in the USA both address this topic by selecting and defining standards to enable flexibility in running workflows composed of meteorological and hydrological models on range of ICT architectures. The DRIHM project goes a step further by developing a distributed computing infrastructure and a related user portal to simplify the running of such workflows on a complex distributed, heterogeneous infrastructure. This is the reason why DRIHM2US based its activities by taking guidelines and services formulated or applied as part of the DRIHM project, harmonizing them with their counterparts in the USA.

As stated in the DRIHM2US Description of Work, the main topics considered are

1. A comparative analysis of certain components underlying the present e-Science Infrastructures (e-Infrastructures) supporting hydro-meteorological research activities both in Europe and the United States, with the goal to present the main technical elements to consider in future designs of interoperable solutions;
2. Assessment of options for the structure and operation of a future organisation: drawing on publicly available information obtained mostly from personal contacts to allow the assembly of a set of requirements and options for any post-project managerial structure

In Section 3 the HMR point of view about the first point is presented, while the ICT guidelines are sketched in Section 4. The second point will be addressed in Section 5, followed by some conclusions and future directions.

3 Domain Expert Networking

In this section the outcome of the planned four domain expert networking meetings, two in Europe and two in the USA, are briefly summarized. Two of the meetings were held at conferences (the large European Geosciences Union general assembly - EGU 2013 and the smaller Congress of the International Environmental Modelling and Software society – iEMSs 2014), the third one was a private project meeting and the fourth alongside the DRIHM 2014 seasonal school. Overall, this combination served its purposes in enabling ideas to be developed and explored, relationships to form and the respective projects to be coordinated.

The potential offered by present e-Infrastructures is clear, as is the current diverse and complex nature of them. Starting from the analysis of some of the ideas devised on the DRIHM project, the main issues arising from the current positions of such infrastructure and how to characterise and overcome the future challenges are presented below.

The core theme is 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. Attributing these responsibilities in a proper and smart way represents the key for the success: more than 10 years' work has gone into getting where we are today and the trends are becoming clear. In particular three main aspects were considered:

- **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. 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; they simply want to buy a better answer;

- **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 base 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. However, 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. Ultimately, model operation through the command line gives optimal control and is necessary anyway for batch running of models.
- **Future Paradigms: 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. Also, controlled vocabularies are necessary and desired. The assumption that they will all use the English language may be disenfranchising non-English speakers.

4 ICT Architecture Harmonization

While today HMR scientists regard computational techniques as a third pillar alongside experiment and theory as already shown by the DRIHMS project, e-Science evolved in the last couple of years as the "global collaboration in key areas of science and the next generation infrastructure that will enable it." This definition includes High Performance Computing (HPC) silos, High Throughput Computing (HTC) clusters, Grids, Clouds, and many more.

Form the ICT point of view, the main goal of DRIHM2US is to sustainably facilitate a more systematic approach to developing, deploying and executing scientific HMR applications across geographies over emerging and future e-infrastructures. Key to the success of the DRIHM2US project is therefore the identification of a set of mechanisms to ensure the sustainability of developed technologies and best practices beyond the lifetime of the project.

Based on the analysis of HMR use cases and the definition of a common architecture model for HMR e-infrastructures several gaps were identified (the sequence does not imply any order of importance):

- Ability to easily access HM data repositories, models and computing resources;
- Facilitation of collaboration between meteorologists, hydrologists and Earth Science experts;
- Easy (and fool-proved) addition of new application components (e.g., models) to existing scientific workflows;
- Standardized graphical user interfaces (GUI) (including virtual reality integration);
- Unified integrated modelling platforms;
- Dynamic integration of different simulation components in a way that will allow more accurate modelling and prediction of hydro meteorological phenomena;
- Integration of privately operated meteorological stations;
- Transparent, secure and consistent access to ICT infrastructures (like Grid portals and science gateways) which have to be tailored to the expected user communities;

- Orchestration of simulation components by coupling appropriate models on various scales using standardized interfaces;
- Execution of designed experiments on resources typically unknown in advance;
- Data brokerage;
- Workflow interoperability solutions to allow workflow sharing;
- Allow HMR models to interact with each other during the execution of the simulation (tight coupling);
- Split large and often unmanageable applications into sets of linkable components which solve singular problems.

An important aspect of future integration of trans-continental e-infrastructures is their potential interoperability (or more precisely: the interoperability of components of such infrastructures). Additional requirements were identified within the consortium meetings and interactions:

- **To form a new role in terms of integrated modelling ('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.
- **To formally distinguish between roles:** 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 these allows model frameworks to be tailored appropriately for these (and any other) valid uses.
- **To integrate more 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.
- **To standardize more and faster:** 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. On the other hand, standardization processes tend to be too slow. We have to find ways for acceleration.

The concept of sustainability matrix was used to approach this challenge. From the matrix we derived requirements and implementation options for a sustainable organisational setup. Methodologically, we applied transformation rules to map the matrix onto organizational specifications. Every matrix cell was assigned one or more structural elements, every such element was associated one or more processes.

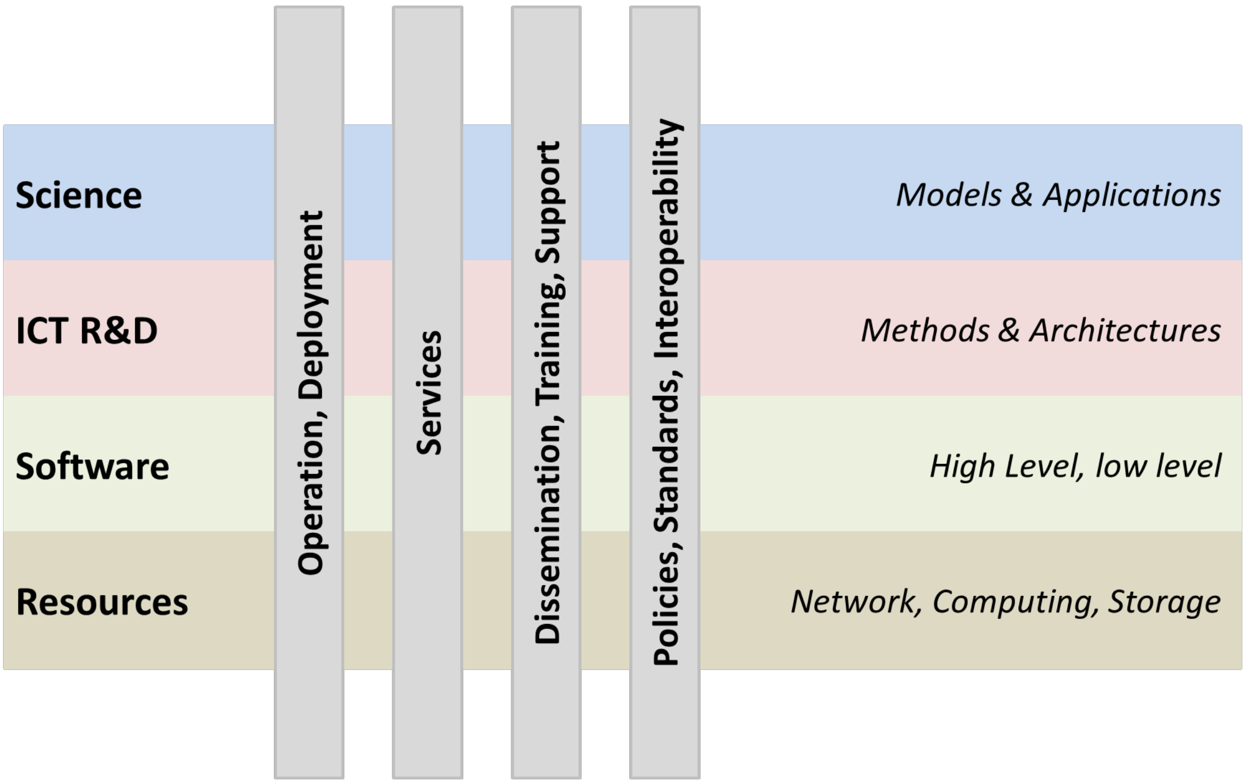

Figure 1 The DRIHM2US sustainability matrix.

Starting from these considerations a future integration plan for technical elements was derived. The goal is to provide guidelines that should be taken into account when designing interoperable and integrated trans-continental infrastructures.

Consequently, short-term objectives, mid-term objectives, and long-term objectives are summarized in the following table as arising from the outcomes of D2.3 Opportunity & gap analysis and D3.3 Domain expert networking report.

Short-term objectives	*to be achieved within the next 1-2 years*
Expected Achievements	• expand the HMR e-Infrastructure community in terms of number of organisations, number of projects, number of applications worldwide and also of application domains, extending the scope to Earth Science at large • reduce the technological gaps by providing a set of formal use cases, a set of data and metadata descriptions, and a set of basic workflows services • "gridify" HMR applications to use standards (initially for dissemination purposes, but gradually moving to production-type applications) • Provide wrappers, bridges, and adapters to solve the most urgent interoperability related issues
Means (examples)	• creation of synergies between various HMR and HMR-related communities and continuous promotion and dissemination activities (e.g., the EGU/AGU assembly conferences) • definition and adaptation of an extensibility concept for the DRIHM portal which includes the integration of US requirements and and an XSEDE plug-in • promoting training activities on regular basis at the international level to attract an increasingly larger amount of users of the HMR services • additional funding sources

Table 1 Short-term objectives.

Mid-term objectives	*to be achieved within the next 2-3 years*

Expected Achievements	• Identification of and solutions to blocking factors for using HMR e-infrastructures, including
	  ○ data access
	  ○ computing resources access
	  ○ quality of service
	• Porting of HMR applications (also from different research disciplines)

Mid-term objectives	*to be achieved within the next 2-3 years*
	○ to use identified standards ○ publication of scientific results which could not have been achieved without such a standardized infrastructure • Facilitate easy access for citizen scientists to HMR data and workflows to get a better understanding of the complexity in predicting HIWE and related impacts • Develop a more abstract wrapper, bridge, adapter concept than short-term
Means (examples)	• increase the number of HMR applications to use standardized services and interfaces and let them benefit immediately for having access to sensors, data, archives, etc. • expose and publish the interfaces and standards • install the basic HMR profile • development of an HMR e-infrastructure aiming at supporting all interoperability patterns (see above) • definition of an enhanced version of the API for data access, integrated single-sign-on, definition of (or contribution to) metadata ontologies • define dynamic authentication and authorization policies for accessing data and workflows to assure science related quality-of-service levels on demand • define methods for easy installation of forecasting chains and their visualization

Table 2 Mid-term objectives.

Long-term objectives	*to be achieved within the next 4-6 years*
Expected Achievements	• Establish a standardized HMR e-infrastructure fully integrating all relevant standards and compliant with the profile(s) defined mid-term ○ The e-infrastructure should facilitate an easy and seamless access to distributed and diverse data and workflows.
Means (examples)	• definition of an interoperability-driven HMR e-infrastructure • make HMR data and workflows readily available via respective infrastructure services • establish within the next 6-8 years a framework to provide timely data of interoperable early warning forecasts for local, national,

Long-term objectives	*to be achieved within the next 4-6 years*
	regional, and international policy makers moving the HMR e-Infrastructure services to a Technological Readiness Level (TLR[1] equal to 8 or higher ○ This requires a capacity building strategy for HMR that will significantly strengthen the capability of all countries to solve HMR related challenges.

Table 3 Long-term objectives.

The implementatio of some of the short-tem objectives has already been started by leveraging the work performed in the DRIHM project (e.g., portal integration, external resource allocation) and the standards based activities of the Research in Advanced Distributed Cyberinfrastructure and Applications Laboratory (RADICAL) initiative led by DRIHM2US partner Rutgers

[1] http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

5 Sustainable International Research Infrastructure

This section focuses on the sustainable maintenance and further development of a future e-Infrastructure developed on the basis of the guidelines presented in the previous Section. This result was obtained starting from the analysis of international organizations and communities such as PRACE, EGI, XSEDE, CSDMS, RDA, and EarthCube.

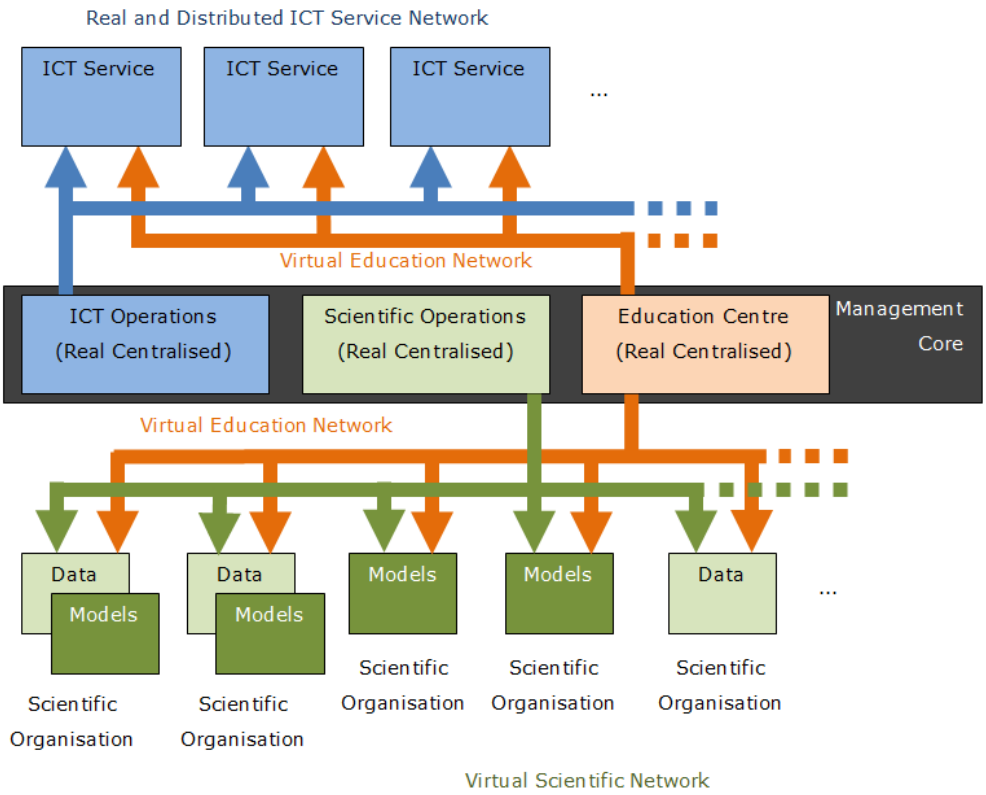

Figure 2: Recommended organizational structure .

The organizational structure is given in Figure 2. It consists of a management board supported by three working groups: ICT operations, scientific operations, and an educational centre. In the current status with a relatively small user community for the infrastructure, we start by implementing the structure as a virtual organization with three working groups consisting of a

subset of the project partners (selected from DRIHM, DRIHM2US, and SCIHM). As the main organization, each working group will be a virtual organization in itself with a single, real institution as the lead partner and working group chair. The management board consists of the chairs of the three working groups.

To enable further growth an international advisory board consisting of strategically chosen representatives from the partner institutes and associated organization such as listed above. As the infrastructure matures extra groups may be defined, and the overall organization may evolve from the initial virtual organization into a real entity. **Errore. L'origine riferimento non è stata trovata.** shows the core organization consisting of the three working groups supporting the management board. The services provide (visualized using the straight thin blue arrows) are offered to all modellers in the community. The next sections will address each group individually.

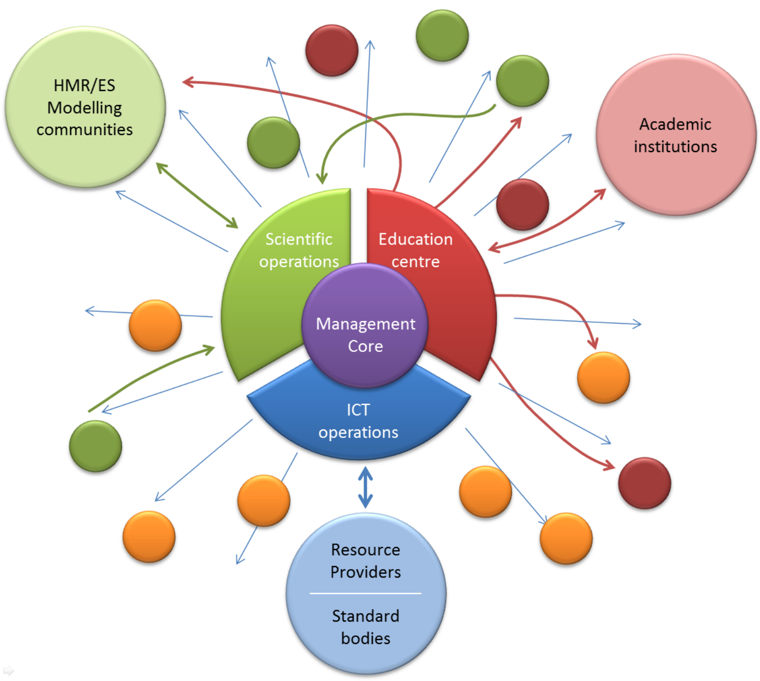

Figure 3: Organizational structure with interactions.

5.1.1 ICT operations

The activities of the ICT operations working group encompasses the updating, maintenance, and management of the ICT services of the infrastructure including proper incident management and provisioning of technical support to users on these services. This working group has strong interaction with the resource providers, developers of innovative core ICT services, and relevant standard bodies. It provides technical testing services to the scientific operations working group, and assist the educational working group with providing training to the community.

5.1.2 Scientific operations

The scientific operations working group focuses on the updating, maintenance, and management of scientific applications and accessibility of relevant data sources on the ICT infrastructure, and provisioning of technical support to users on these services (such as help with porting applications, and effective use of resources by code developers and end users). This group interacts strongly with the scientific community, especially the scientific modelling community, to determine a strategy for models to be added to the infrastructure. The green arrows represent the incorporation of new models in the infrastructure. In collaboration with the scientific community it verifies the scientifically correct implementation of new applications on the infrastructure before public release. This working group prepares the applications for the testing and deployment by the ICT operations working group; the interactions of the applications with each other, and with the infrastructure are based on standard. The group interacts with the educational group to reach out to the community at conferences and curricula to teach about the standards and applications deployed.

5.1.3 Education centre

The educational working group develops in collaboration with the other two working groups material for infrastructure training events, workshops, and curriculum activities at academic

institutions. The intended audience includes students of the relevant HMR and Earth Sciences domains, ICT students, and staff of research institutes, consultants, and operational centres. The red arrows represent the training of modellers and developers on using and contributing to the infrastructure. Special training events may be organized at popular modelling conferences.

6 Conclusion

Based on the previous discussion, it may be concluded the future of e-Infrastructures for Earth Science research will be largely driven by the need of finding a closer and more stable cooperation with operational Institutions dealing with the predictions and the preventions of HIWE and their impacts, in order to ensure a real translation of the innovative research results produced by such e-Infrastructures into operational tools and services. The project DRIHMS, with its whitepaper, DRIHM with the related e-Science environment, and DRIHM2US with its interplay with US cyber-infrastructures already paved the way towards this very important goal.