e-infrastructure

CAPACITIES

DRIHM2US

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

D5.2: Assessment of Options for an Organizational Setup

Abstract: In the context of a sustainable international research infrastructure this document assesses options for an organizational setup.

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

cima RESEARCH FOUNDATION · LMU MNM · imati · HR Wallingford · Deltares · CUAHSI · NCAR · RUTGERS

Document Information Page

Contract Number	313122
Project Name	Distributed Research Infrastructure for Hydro-Meteorology to United States of America
Project Acronym	DRIHM2US
Deliverable Number	5.2
Deliverable Name	Report on an Assessment of Options for Organizational Setup
Work Package Number	5
Work Package Name	Sustainable International Research Infrastructure
Deadline	30/09/2013
Version	1.0
Dissemination Level	PU
Nature	R
Lead Beneficiary	LMU

Document History

Date	Version	Description
6th Sept 2013	0.8	Initial write up for review by the project team.
24th Oct 2013	0.9	Reviewer's comments integrated
16th Nov 2013	1.0	Finalized version

Table of Contents

List of Figures

List of Tables

1 Executive Summary

The objectives of Work Package (WP) 5 (Sustainable International Research Infrastructure) are related to the sustainability of international research infrastructures. As part of the project's activities a first set of prototypes for an EU/US HMR[1] Science platform will be specified. À la longue they may serve as the building blocks for a global interdisciplinary research platform. The first objective of this WP will therefore be to elicit research infrastructure needs, and to identify opportunities in fulfilling these needs.

Although the project's focus is on EU and USA, the outcome is of global importance. With an international research infrastructure as a vision, the question emerges how such a research infrastructure can be maintained, operated, and improved over time. Critical aspects of joined operations concern governance and legal issues (liabilities, ownership, applicable law, accessibility, openness) and, of course, funding of projects aiming at using and advancing the infrastructure. Only if these issues are satisfactorily dealt with the required credibility will be achieved to attract users and investors. The second objective of this work package is therefore to set up a post project organization for maintaining and for further developing the platforms/e-science environments to sustainably facilitate innovative HMR Science.

This deliverable reports on the results of task 5.2 which investigates the options for such an organizational setup. The findings are based on the proposal of a sustainability matrix which relates a science specific perspective, an ICT perspective, a software perspective and a resource perspective to operational, service, dissemination and policy aspects. A set of transformation rules is used to derive organizational specifications which eventually lead to options for an organisational setup. We will discuss three options: a) the centrally implemented real organization paradigm; b) the distributed implemented real organization paradigm; c) the

[1] More generally speaking, the focus is on Earth Sciences. However, in this report we use Earth Sciences and Hydro-Meteorological Sciences interchangeably.

virtual organization paradigm. Based on these options we recommend an organizational setup as a mix of all.

2 Introduction

2.1 Background and Approach

The objectives of Work Package (WP) 5 (Sustainable International Research Infrastructure) are related to the sustainability of an international research infrastructure, in concreto to the building blocks of a sustainable EU/US Earth Science platform. The first objective of this WP will therefore be to elicit and prioritize research infrastructure needs, and to identify opportunities in fulfilling these needs. The second objective is to set up a post project organization for maintaining and for further developing the platform/e-science environments. This report is devoted to the latter task, while deliverable D5.1 investigates the former one in more detail. However, all reports generated in WP5 should be considered to form a unit as all of them cover both organizational aspects and technology ones.

A fundamental prerequisite when discussing organizational setups is a thorough understanding of the visionary goals to achieve. These goals are not only related to various fields of interest (like pure science or resources), but also to various orthogonal perspectives (e.g., standardization or service management or training). The goals are reflected in the DRIHM2US sustainability matrix to be discussed in section 3.

With the matrix in mind we can define a set of transformation rules to transform the matrix cells into structural and process requirements which then lead to organizational specifications implying setup options. This will be done in section 4.

A concluding remark will be given in section 5.

2.2 Remark

This report is part of the deliverables of WP5. Although the primary goal of the report is an assessment of options for organizational setups, it is significantly intertwined with other activities in WP5:

- With deliverable D5.1 (Overview, assessment, plan and prioritisation of development needs) because it not only provides a framework for this assessment, it also sets the context for the investigations in 5.1 and it presents some of the high-level requirements.
- With deliverable D5.3 (Outline of organizational structure) because it sets the scene for these solicitations.
- With deliverable D5.4 (Detailed organization implementation plan) as it presents some of the mandatory buildings blocks to be agreed upon.

3 What Needs to be Sustained?

The declared goal of DRIHM2US is to facilitate a more systematic approach to developing, deploying and executing scientific HMR applications across geographies over emerging and future e-infrastructures (which by the way are also shared by other research communities). Key to the success of the DRIHM2US project is therefore a set of mechanisms to ensure the sustainability of developed technologies and procedures beyond the lifetime of the project. The general DRIHM2US sustainability approach is exhibited by the sustainability matrix of Figure 1 and the dimensions the matrix is spanned by.

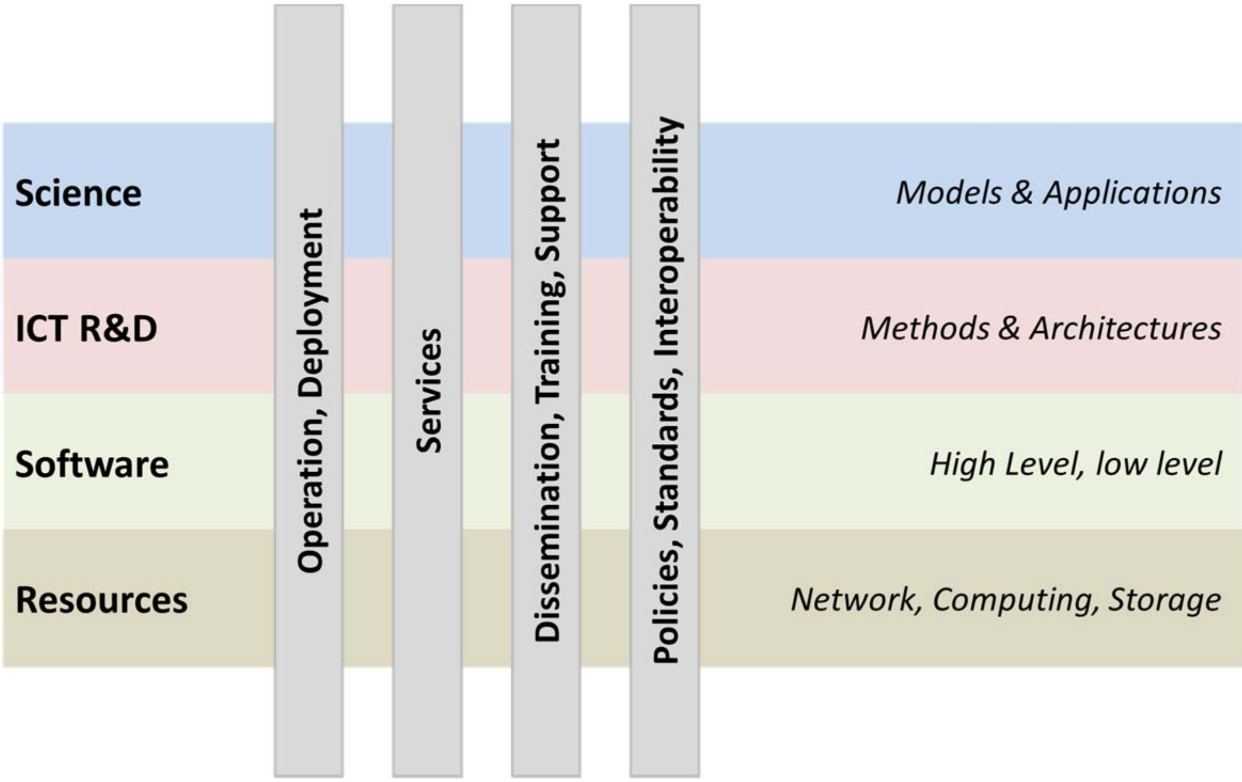

Figure 1: DRIHM2US Sustainability Matrix

The goal of the matrix is to provide a methodologically sound classification scheme for a subsequent assessment of options for an organizational setup. The matrix rows indicate

different aspects of science and engineering, both in computer science (ICT R&D, software, resources) and in HMR (science). The columns indicate tasks and activities to support these aspects.

The DRIHM2US Sustainability Matrix has been derived by analysing the e-IRG Roadmaps 2010 [4] and 2011 [6] and by assessing different e-infrastructure frameworks for HMR in Europe and in the USA [2]. It follows an approach similar to the ones proposed in [4] and [6] by considering the following layers:

- *Resources* (network resources, computational resources, storage resources),
- *Software* with low-level basic services and high-level compound services that draw on components of a resource layer,
- *ICT Research and Development* comprising methods and architectures for facilitating HMR
- *Science* devoted to scientific HMR and Earth Science applications

3.1.1 The Science Layer

The majority of HMR applications use models and data from various sources in order to execute workflows as exemplarily shown in Figure 2 using the DRIHM project[2].

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

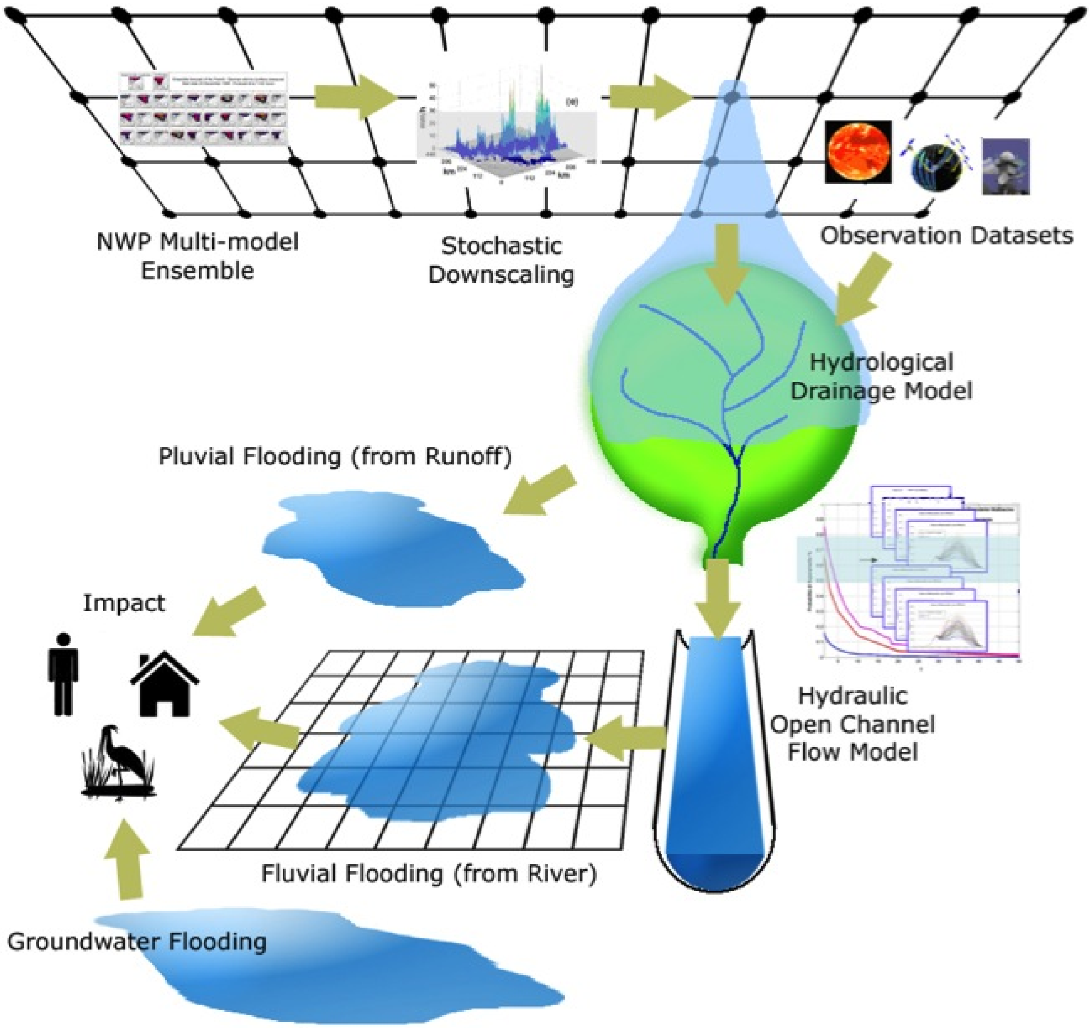

Figure 2: Typical HMR Workflows (DRIHM Example) [2]

Hence, a key part of DRIHM2US's approach to sustainability is to ensure that HMR concepts and solutions are used by a wide range of science and engineering users that require model coupling and forecasting capabilities. Initially, DRIHM2US is focusing on a small set of application areas like the ones shown in Figure 2. Consequently, the principal approach of the science layer consists of facilitating support and training related to the deployment and usage of HMR tools for scientific investigations with the EU and US HMR and related Earth Science communities.

An important aspect to consider on this layer is the evolution of European policies for data access towards the approaches already implemented in USA, which exhibit a consistent open data policy that forms the legal framework for public accessibility of huge amounts of federal data. As technology progress and the development of standards advances, we will see more and more of these data is available online[3]. A large part of hydro-meteorological data in the US is collected by federal agencies and therefore freely available. The DRIHM2US partner CUAHSI[4] represents a huge community that has experience in harvesting the hydro-meteorological data from various resources and making them available for research.

Sustainability on the science layer can be supported by attracting funding for new research projects in these areas, possibly through further joint European Commission and National Science Foundation funding mechanisms. An additional element of sustainability on this layer is concerned with the development of new curricula teaching HMR related e-science, in particular with emphasis on model coupling and data sharing. Standardization on this layer is work-on-progress and is mainly reported in deliverable D2.3 [7]. Finally the sustainability on this layer can be also promoted through a proper collection of operational questions and requirements from public servant institutions, like Civil Protection and Disaster Risk Reduction agencies, as well as insurance and reinsurance companies[5].

Typical examples of sustainability activities (taken from the FP7 DRIHM project) are summer schools, publications and presentations at conferences and workshops. However, sustainability is not only focused on effective dissemination and training, but also on building a sustainable community for HMR. For example, by contributing to the European Geosciences Union (EGU)[6]

---

[3] The website http://catalog.data.gov/dataset is a portal that provides an overview of which data sets are available somewhere.

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

[5] E.g., Munich Re http://www.munichre.com/de/homepage/default.aspx

[6] http://www.egu.eu/

and to the American Geophysical Union (AGU)[7] community effects may be fostered to further increase the public awareness of HMR and their potential bolstering of innovative disaster management.

3.1.2 The ICT R&D Layer

Sustainability on the ICT R&D layer concerns the ICT scientific methods and architectures per se and especially those that enable HMR. The outputs of these activities will be sustained by disseminating results as for example generated by projects and initiatives like DRIHM, SCIHM[8] and CUAHSI[9] to a wider community. This needs to be accomplished by dedicated training activities, adequate technical documentation, and high-quality (ICT) publications for peer-reviewed scientific journals and conferences. Another more technical example is the provisioning of automatic architecture configurators. Also on this layer is sustainability not only focused on dissemination and training, but also on building a sustainable collaboration with infrastructure providers like EGI[10], PRACE[11] and XSEDE[12].

---

[7] http://sites.agu.org/

[8] http://scihm.org/

[9] http://his.cuahsi.org

[10] http://www.egi.eu/

[11] http://www.prace-ri.eu/?lang=en

[12] https://www.xsede.org/

3.1.3 The Software Layer

The process of creating HMR software (services, applications, models) and integrating it with existing e-infrastructures is based on predefined reusable software components. As pointed out in [2], several project consortia have been/are/will be contributing a number of software components that need to be integrated with and maintained by e-infrastructures to provide new or improved capabilities for scientific users.

From the sustainability perspective, it is important to understand two different software development approaches. In order to ensure early adaptation, integration and deployment of a minimal set of infrastructure components to facilitate the execution of HMR applications, a fast development may be suitable. A more thorough development cycle, on the other hand, may require more formal adaption, integration and deployment of higher level services in order to realise a fully integrated HMR infrastructure. This distinction has for example been realized in the EU-funded MAPPER project[13] with their “fast-track” and “deep-track” approach. The second approach is more relevant to sustainability as it goes a step further and provides automation mechanisms based on reusability and standards. “Reusable components” in this context is not only code, but also software specifications, reference implementations of standards, application use cases or workflow templates.

From a sustainability and maintenance perspective, HMR software components fall into four groups (see also [2]):

- Local domain (low level) services like middleware services, hosting environments or workflow engines
- Infrastructure domain (high level, e.g. Grids) services and user interfaces like portals that facilitate application specific services to utilize infrastructure resources.
- Common community independent services like data conversion libraries

[13] http://www.mapper-project.eu

- HMR specific services like model couplings tools based on the adoption of specific standards for data staging in and staging out (e.g NetCDF Climate and Forecast (CF) Metadata Convention).

Similarly to resources, sustainability of software also needs to be considered from various angles: from an operational perspective, a service perspective, an educational perspective, and an interoperability perspective.

3.1.4 The Resource Layer

Resources in the broader sense of this deliverable are physical resources like network elements, computing elements, storage elements, instruments and sensors. Such resources are "owned" by resource providers (i.e., organisations which are legal entities) and provided to Virtual Organisation (VO) for temporary use, subject to policies and constraints. Consequently, the perceived (virtual) life cycle of resources is coupled to the life cycle of VOs and VO membership of resource users. After a VO is decommissioned, the resources are no longer available to the members, unless they join another VO that grants new access rights. VOs thus represent structures implemented over e-infrastructures that facilitate access to resources that span multiple administrative and geographical domains.

HMR applications inherently rely on a variety of resources from different providers for various VOs. Examples in Europe include HPC systems provided through the Partnership for Advanced Computing in Europe (PRACE)[14], the data communications infrastructure provided by DANTE[15], and the storage capabilities and High-throughput Computing (HTC) services available through EGI and the NGIs. Examples in USA are provided by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by the National Science Foundation (NSF), which

[14] http://www.prace-ri.eu/?lang=en

[15] http://www.dante.net/

has replaced and expanded on the TeraGrid project. However a major issue arises when comparing the resource access policies for EGI and PRACE: while EGI grants the access to its computational sites to the members of a VO, possibly with the use of robot certificates, the PRACE policy is to create personal accounts corresponding to only the partners that presented the peer reviewed HPC research proposals. It is therefore clear that it may be burdensome for non-partner of a given proposal to run high-resolution simulations on this infrastructure. Because HMR – as a community – is not designed to provide any resources to VOs, HMR is completely decoupled from resource provisioning by respective resource providers. On the other hand, the intention of HMR related projects is to sustainably provide HMR services. Consequently, arrangements have to be made on the resource level for enabling long-term access to required resources, both for members of VOs whose life span outlives current HMR projects and for members of VOs whose establishment is scheduled for the future.

Resource sustainability needs to be considered from different angles: from an operational perspective, a service perspective, an educational perspective, and an interoperability perspective. Assessing the sustainability of resources in the operational/deployment view requires differentiating between resources provided by “independent” resource providers (i.e., those who do not contribute to HMR projects directly) and resources provided by the HMR community. Strictly speaking, sustaining any resources (of independent resource providers and community resources) is thus beyond any HMR project’s scope. Despite this general restriction, however, there is a difference between independent resources and community provided ones. While the life cycles of the former are completely independent from HMR activities, the life cycles of the latter may have a relationship with HMR activities, which is typically externalized as

1. a direct project contribution (e.g., parts of the cluster at the Leibniz Supercomputing Centre);
2. a Service Level Agreement (SLA) formally specifying the terms and conditions of guaranteed resource provisioning (this option may apply even for the adoption of the infrastructure for production purposes – i.e. agreement with a cloud provider);

3. a Memorandum of Understanding (MoU) describing an agreement with the community sites (acting as resource providers) for a convergence of will with an intended common line of action related to resource provisioning;
4. execution of production simulation on dedicated resources (this option is limited to adoption of the infrastructure for production purposes)

Note that most of these externalizations are time-dependent: While a direct contribution (bullet 1) is only possible during the lifetime of the project, all other externalizations (bullets 2, 3 and 4) may also be relevant after a project's end.

Sustainability (of resources) would thus benefit from MoUs and SLAs with resource and infrastructure providers for long-term and dependable resource provisioning not only in Europe (PRACE, EGI) but also in the USA (XSEDE[16], ExTENCI[17]) and other parts in the world.

Sustainably allocating, using and managing resources requires dedicated standards and demonstrated compliance with these standards. This is the background of proposing in this context an HMR Basic Profile (see the appendix in section 7).

[16] https://www.xsede.org/

[17] https://sites.google.com/site/extenci/

4 Requirements and Options for an Organizational Setup

In this section we take the sustainability matrix of section 3 to derive requirements and implementation options for a sustainable organizational setup. Methodologically, we apply a set of rules $\mathcal{R}$ to transform the sustainability matrix $\mathcal{M}$ (see Figure 1) into an organizational specification $\mathcal{O}$, possibly adhering to constraints of a set $\mathcal{C}$.

4.1 Transformation Rules

The transformation rules map the sustainability matrix $\mathcal{M}$ onto an organizational specification $\mathcal{O}$ in that they assign every matrix cell one or more structural elements and one or more processes associated to these elements[18]. In concreto, the matrix consists of 16 basic cells[19], defined by the junctions of row and columns in Figure 1:

Table 1: Cells of the Sustainability Matrix

Cell	Description
$\mathcal{M}_{\text{sciene,operation}}$	This cell covers all structural elements and processes on and between them referring to *operational* aspects of HMR specific activities.
$\mathcal{M}_{\text{sciene,service}}$	This cell covers all structural elements and processes on and between them referring to the definition, delivery of and access to HMR specific

[18] For the purpose of this report this is sufficient. A formal organization theory, however, needs to go well beyond the scope of this deliverable.

[19] We omit for simplicity a more detailed view on the cells. Formally speaking, the cells themselves represent sequences (often hierarchies) of more fine-grained cells.

Cell	Description
	services.
$\mathcal{M}_{sciene,DTS}$[20]	This cell covers all structural elements and processes on and between them referring to the dissemination of, the training in, and the support of HMR specific science activities.
$\mathcal{M}_{sciene,PSI}$[21]	This cell covers all structural elements and processes on and between them referring to policy aspects, the standardization of, and the interoperability of HMR specific science activities.
$\mathcal{M}_{ICT,operation}$	This cell covers all structural elements and processes on and between them referring to operational aspects of computer science specific activities.
$\mathcal{M}_{ICT,service}$	This cell covers all structural elements and processes on and between them referring to the definition, delivery of and access to computer science specific services.
$\mathcal{M}_{ICT,DTS}$	This cell covers all structural elements and processes on and between them referring to the dissemination of, the training in, and the support of computer science specific activities.
$\mathcal{M}_{ICT,PSI}$	This cell covers all structural elements and processes on and between them referring to policy aspects, the standardization of, and the interoperability of computer science specific activities.
$\mathcal{M}_{software,operation}$	This cell covers all structural elements and processes on and between them referring to operational aspects of software specific activities.
$\mathcal{M}_{software,service}$	This cell covers all structural elements and processes on and between them referring to the definition and delivery of software specific services.
$\mathcal{M}_{software,DTS}$	This cell covers all structural elements and processes on and between

[20] DTS = Dissemination, Training, Support

[21] PSI = Policies, Standards, Interoperability

Cell	Description
	them referring to the dissemination of, the training in, and the support of software specific activities.
$\mathcal{M}_{software,PSI}$	This cell covers all structural elements and processes on and between them referring to policy aspects, the standardization of, and the interoperability of software specific activities.
$\mathcal{M}_{resources,operation}$	This cell covers all structural elements and processes on and between them referring to operational aspects of resource specific activities.
$\mathcal{M}_{resources,service}$	This cell covers all structural elements and processes on and between them referring to the definition and delivery of resource specific services.
$\mathcal{M}_{resources,DTS}$	This cell covers all structural elements and processes on and between them referring to the dissemination of, the training in, and the support of resource specific activities.
$\mathcal{M}_{resources,PSI}$	This cell covers all structural elements and processes on and between them referring to policy aspects, the standardization of, and the interoperability of resource specific activities.

The cells can now be transformed into requirements regarding the structure of an organizational setup and the respective processes by something we call "Transformation Rule" in this report.

Table 2: Transformation Rules

Id	Cell	Structure Requirement	Process Requirement
R1	$\mathcal{M}_{sciene,operation}$	• A group that deploys and supports HMR applications on various production infrastructures • A group that liaises with HMR application providers and infrastructure providers	• Processes to deploy and support HMR applications on various production infrastructures
R2	$\mathcal{M}_{sciene,service}$	• A group that deploys and supports technical tools and services that are directly related to HMR on various	• Processes to deploy and support technical tools and services on various production infrastructures

Id	Cell	Structure Requirement	Process Requirement
		production infrastructures	
R3	$\mathcal{M}_{\text{sciene,DTS}}$	• A group that provides HMR specific trainings, curricula and usage support • A group that creates HMR specific scientific publications	• Processes to deliver and monitor HMR specific trainings, curricula and usage support • Processes to push HMR specific scientific publications
R4	$\mathcal{M}_{\text{sciene,PSI}}$	• A group that liaises with HMR specific standardization bodies • A group that liaises with HMR specific funding authorities	• Processes to support HMR specific standardization processes • Processes to monitor and trigger the application to HMR specific funding calls
R5	$\mathcal{M}_{\text{ICT,operation}}$	• A group that cooperates with ICT service providers for deploying ICT services on various production infrastructures	• Processes to cooperate with ICT service providers for various production infrastructures
R6	$\mathcal{M}_{\text{ICT,service}}$	• A group that deploys and supports ICT system services on production infrastructures	• Processes to deploy and support ICT system services on production infrastructures
R7	$\mathcal{M}_{\text{ICT,DTS}}$	• A group that provides ICT specific trainings, curricula and usage support • A group that creates ICT specific research publications	• Processes to deliver and monitor ICT specific trainings, curricula and usage support • Processes to push ICT specific scientific publications
R8	$\mathcal{M}_{\text{ICT,PSI}}$	• A group that liaises with ICT specific standardization bodies • A group that liaises with ICT specific funding authorities	• Processes to support ICT specific standardization efforts • Processes to monitor and trigger the application to ICT specific funding calls
R9	$\mathcal{M}_{\text{software,operation}}$	• A group that collects, evaluates, deploys and supports software packages for various production infrastructures	• Processes that collect, evaluate, deploy and support software packages for production infrastructures
R10	$\mathcal{M}_{\text{software,service}}$	• A group that identifies software packages to be sustained by various communities	• Processes to identify software packages to be sustained by communities
R11	$\mathcal{M}_{\text{software,DTS}}$	• A group that provides or contributes to special software package trainings • A group that provides special software support	• Processes to support special software package trainings • Processes to support special software
R12	$\mathcal{M}_{\text{software,PSI}}$	• A group that liaises with standardization bodies • A group that liaises with	• Processes to support the development of the HMR Basic Profile

Id	Cell	Structure Requirement	Process Requirement
		infrastructure providers and service providers	• Processes to grant compliance with the HMR Basic Profile
R13	$\mathcal{M}_{\text{resources,operation}}$	• A group that provides statistics on how resources are used for HMR	• Processes to monitor HMR resource usage
R14	$\mathcal{M}_{\text{resources,service}}$	• A group that provides statistics on how services are used for HMR on specific resources	• Processes to monitor HMR service usage on resources
R15	$\mathcal{M}_{\text{resources,DTS}}$	• A group that collects statistics and publishes papers on it	• Processes to collect statistics
R16	$\mathcal{M}_{\text{resources,PSI}}$	• A group that provides policies for HMR resource usage • A group that provides SLA templates • A group that collects Best Practices	• Processes to sign and monitor SLAs • Processes to collect and evaluate Best Practices

4.2 Requirements and Options for an Organizational Setup

Following the discussions in section 3 and the transformations identified in section 4.1 we can now specify (some of) the requirements and options for an organizational setup after the project's lifetime. In the following table the identifiers "Ox" actually reflect organizational specifications in $\mathcal{O}$. If necessary we also mention constraints "Cx" of $\mathcal{C}$.

Table 3: Organizational Specifications

Id	Organizational Specification	Crossref to Rules	Constraint
O1	• A group that receives science applications, evaluates them, releases them, deploys them on production infrastructures	R1, R2, R9, R10, R12, R16	Maybe need to differentiate by infrastructure
O2	• A group that receives IT services, evaluates them, releases them, deploys them on production infrastructures	R5, R6, R9, R10, R12, R16	Maybe need to differentiate by infrastructure
O3	• A group that provides trainings	R3, R11, R12, R16	
O4	• A group that provides hands-on support	R3, R11, R12,	

Id	Organizational Specification	Crossref to Rules	Constraint
		R13, R14, R16	
O5	• A group that provides incident management	R12, R13, R14, R16	
O6	• A group that provides workshops on conferences and that solicits for publications	R3, R7, R13, R14, R15	
O7	• A group that defines and deploys student curricula	R3, R7	
O8	• A group that liaises with standardization bodies	R4, R8, R12, R16	
O9	• A group that prepares joint funding proposals	R4, R8	

These organizational specifications have both a hidden structure setup and a hidden process texture. Generally speaking, from a structure perspective any organization consists of "entities" which perform activities directed towards the achievement of commonly agreed organizational aims or objectives. Depending on these objectives, organizations can be structured in many different ways to express an allocation of responsibilities for different functions and processes across these entities. The structure thus determines which entities need to (and hopefully will do) participate in which decision-making processes.

In this report, however, we do not want to prescribe any arrangement of structural entities and processes between them. Rather, we will elaborate on some options for an organizational setup to implement the specifications O1 to O9 of Table 3. There is of course a large number of options. Because it is not our intention to discuss all of them, we select four of them. They are arranged in the following matrix.

	Real Organization	**Virtual Organization**
Central Organization	Option CR	Option VO
Distributed Organization	Option DR	

We will discuss them briefly.

4.2.1 Option CR

This option is characterized by a centrally organized real organization. By a real organization we understand any kind of legal entity, be it economically oriented, an academia setup or a non-profit organization. The main differentiators between virtual organizations and real ones are the (principally) general purpose establishment, the unrestricted lifetime and a single administration (in its broadest sense) of the latter.

This option aims at a classic hierarchy based organization [7] with formally establishing a legal entity to manage the HMR e-infrastructure “eco-system”. The organizational specifications can then be translated into an organization structure like the one depicted in Figure 3.

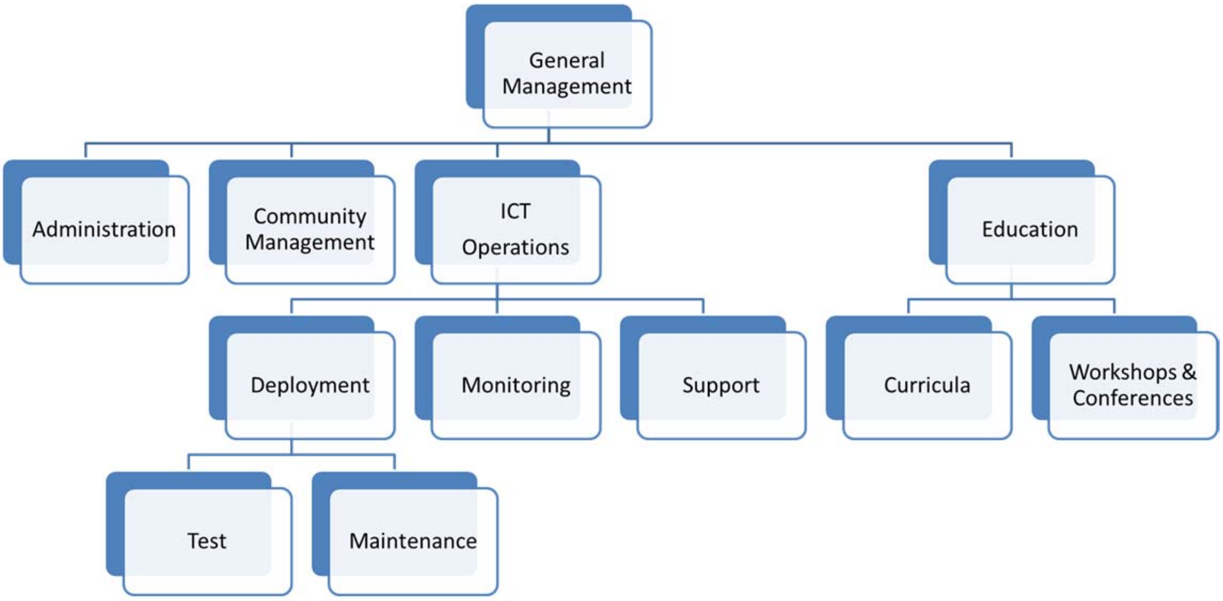

Figure 3: Example of an Organization Structure for Option CR

The implementation of the organizational specifications O1 to O9 of Table 3 is obvious. This option has advantages and disadvantages. Some of them are given in the following table:

Advantages	Disadvantages
• Central decision making • Economic setup • Could serve as sustainability platform for other projects as well (business model)	• Need to establish a legal entity • Legal entity overhead • Liabilities • Intellectual property rights • Licensing

4.2.2 Option DR

This option is characterized by a distributed real organization as for example found for large corporations. It typically aims at a classic matrix oriented organization [7] with formally establishing a legal entity. For this option, the organizational specifications O1 to O9 can be translated into an organization structure like the one depicted in Figure 3, however, with some adjustments.

- The General Management requires more coordination overhead.
- The other "branches" are managed and performed by selected and appropriate specialized "departments" in close cooperation with other ones.

The implementation of the organizational specifications O1 to O9 of Table 3 is again obvious. This option also has advantages and disadvantages. Some of them are given in the following table:

Advantages	Disadvantages
• Central decision making • Economic setup • Could serve as sustainability platform for other projects as well (business model)	• Need to establish a legal entity • Legal entity overhead • Liabilities • Intellectual property rights • Licensing • Cooperation overhead

4.2.3 Option VO

This option is characterized by the concept of virtual organizations. By a virtual organization we understand any kind of collection of people and/or resources of various real organizations collaborating for a limited time to achieve a commonly agreed goal. Typical examples are the well-known virtual organizations (VO) in Grids [3] and the virtual research communities over European e-infrastructures [6]. A virtual organization can be centrally organized or may completely act in a distributed way. The standard setup is the latter one. Please note, however, that virtual organization span multiple and heterogeneous domains, technically and organizationally.

Other than in the previous cases, this option is not aiming at formally establishing a legal entity to manage the HMR e-infrastructure “eco-system”. Rather, it favors a virtual (project management type like) setup without formally fixing legal contracts.

One possible way of implementing the organizational specifications can then be the following:

- The General Management reduces to coordination functions.
- The other “branches” are managed and performed by selected and appropriate legal entities (i.e., real organizations).
- Memoranda of Understanding (MoU) drive the collaboration by gluing structural elements and processes.
- Typically, penalties do not apply. Economic failures will not harm.
- The organization requires a well-defined “shut-down” criterion.
- The organization has no fix “department”. Rather, is allows for dynamically changing teams.

The implementation of the organizational specifications O1 to O9 of Table 3 is obvious for this option as well. Some of the advantages and disadvantages are given in the following table:

Advantages	Disadvantages
• Real expertise drives the building blocks • Dynamically changing constellations • Defined termination	• No real enforcement • Requires SLAs and/or MoUs • Coordination overhead • Liabilities • Intellectual property rights • Licensing

4.3 Recommendation

It should be noted that these option are not applied in a strict manner. Rather, similar efforts in other projects exhibit an appropriate mix of approaches. Examples are (we refer to the home pages of these projects):

- the numerous standardization bodies relevant for HMR which range from ICT related ones (OGF[22], OASIS[23], IETF[24], DMTF[25] – just to name a few) via software engineering related ones (e.g. OMG[26], WfMC[27] or ISO) to HMR specific ones like OpenMI[28], OGC[29], iEMSs[30] or WMO[31]
- community specific collaboration platforms as identified in [2]
- the MAPPER project[32]

---

22 http://www.gridforum.org/

23 https://www.oasis-open.org/standards

24 http://www.ietf.org/

25 http://www.dmtf.org/

26 http://www.omg.org/

27 http://www.wfmc.org/

28 http://www.openmi.org/

29 http://www.opengeospatial.org

30 http://www.iemss.org/society/

31 http://www.wmo.int/pages/governance/policy/tech_regu_en.html

32 http://www.mapper-project.eu

- the Initiative for Globus (IGE) in Europe project[33]

Taking these and several other approaches into account, we propose the following (while referring to Figure 3 for the nomenclature):

Recommendation 1:	Use Option VO as the major guideline.
Recommendation 2:	Outsource the "ICT Operation" to Resource Providers under an Option CR umbrella.
Recommendation 3:	Run the "Education" branch under an Option DR umbrella.
Recommendation 4:	Assign duties to all participants in the VO option to support other fields of interest. Sign appropriate MoUs and SLAs.
Recommendation 5:	Assign special task forces (as a "second level VO option") to crowd additional funding and define appropriate bylaws.

An adequate organizational setup needs to define the organizational structures and processes that ensure that the HMR infrastructure sustains and extends the community's strategies and objectives. The process focus of this approach requires an adequate process model. Respective control objectives and efficient and effective organisational structures, all designed to provide reasonable assurance that HMR science objectives are achieved and that undesired events are prevented or detected and corrected.

[33] http://www.egcf.eu/

5 Conclusion

This document derived a set of options for an organizational setup after the end of the project to sustainably support long-term activities in establishing an Earth Sciences research infrastructure.

We proposed a sustainability matrix which relates a science specific perspective, an ICT perspective, a software perspective and a resource perspective to operational, service, dissemination and policy aspects. A set of transformation rules was then used to derive organizational specifications which eventually led to options for an organisational setup. We discussed three options: a) the centrally implemented real organization paradigm; b) the distributed implemented real organization paradigm; c) the virtual organization paradigm. Based on these options we recommended an organizational setup as a mix of all.

6 Acronyms and References

6.1 Acronyms and Abbreviations

Acronym / Abbreviation	Definition
AGU	American Geophysical Union
CUASHI	Consortium of Universities for the Advancement of Hydrologic Science
DMTF	Distributed Management Task Force
DRIHM	Distributed Research Infrastructure for Hydro-Meteorology
DRIHM2US	Distributed Research Infrastructure for Hydro-Meteorology to United State of America
DTS	Dissemination, Training, Support
EGCF	European Globus Community Forum
EGI	European Grid Infrastructure
EGU	European Geosciences Union
ExTENCI	Extending Science Through Enhanced National CyberInfrastructure
HBC	HMR Base Case
HBP	HMR Basic Profile
HMR	Hydro-Meteorological Research
HPC	High Performance Computing
HPC	High Performance Computing
HTP	High Throughput Computing
ICT	Information and Communications Technology
iEMSs	International Environmental Modelling and Software Society
IETF	Internet Engineering Task Force

IGE	Initiative for Globus in Europe
ISO	International Organization for Standardization
MAPPER	Multiscale Applications on European e-Infrastructures
MoU	Memorandum of Understanding
NetCDF	Network Common Data Form
NetCDF CF	NetCDF Climate and Forecast
NSF	National Science Foundation
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
OGF	Open Grid Forum
OMG	Object Management Group
OpenMI	Open Modelling Interface
PRACE	Partnership for Advanced Computing in Europe
PSI	Policies, Standardization, Interoperability
R & D	Research and Development
SCIHM	Standards-Based Cyberinfrastructure for Hydrometeorology
SLA	Service Level Agreement
VO	Virtual Organization
WaterML	Water Markup Language
WfMC	Workflow Management Coalition
WMO	World Meteorological Organization
XSEDE	Extreme Science and Engineering Discovery Environment

6.2 References

[1] Parodi, A. et al.: DRIHM2US Description of Work (DoW), 2012.

[2] DRIHM2US: Report on an Assessment of Current e-Infrastructure Approaches for Hydro-Meteo Research in Europe and the US, Report D2.1 of the DRIHM2US Project, 2013

[3] Foster, Ian ; Kesselman, Carl ; Tuecke, Steven: The Anatomy of the Grid: Enabling Scalable Virtual Organizations. In: International Journal of High Performance Computing Applications 15 (2001), Nr. 3, S. 200-222

[4] Rosette Vandenbroucke: e-IRG White Paper 2011; 2011, http://www.e-irg.eu/images/stories/e_irg_whitepaper_and_comments_2011.zip

[5] Mark Nottingham, Claus von Riegen: Conformance Claim Attachment Mechanisms Version 1.0, http://www.ws-i.org/Profiles/ConformanceClaims-1.0-2004-11-15.html, 2004

[6] e-Infrastructure Reflection Group (e-IRG): e-IRG Roadmap 2010. www.e-irg.eu/images/stories/eirg_roadmap_2010_layout_final.pdf

[7] Camarinha-Matos, Luis; Afsarmanesh, Hamideh; Ollus, M.: *Virtual organizations: Systems and practices.* New York: Springer, 2005

7 Appendix: Outline of an HMR Basic Profile

The requirements for an HMR Basic Profile (HBP) are derived from two main HMR scenarios: from the forecasting chains as described by the DRIHM project (see also Figure 2); and from the underlying e-infrastructure context as described in [2] and provided by e.g. PRACE, XSEDE, EGI. Both scenarios have in common the co-allocation of heterogeneous resources hosted by providers, the model coupling facilitation, and the interoperability requirement. From the point of view of a single service provider, however, the two scenarios reduce to a simpler HMR Base Case (HBC).

The simplest HMR use case is a single HMR model running as a job on a single Grid Computing element (which may be a high performance computing (HPC) system). What distinguishes HBC from a pure vanilla batch job submission use case is that

- the job must be submitted into an advance reservation context,
- the computing element must support parallel jobs as most of the HMR applications require HPC resources.

The users may belong to several autonomous organizations but belonging to the same Virtual Organization (VO), an inherent Grid requirement.

A typical HMR application consists of (see also Figure 4):

- software modules simulating certain phenomena in certain time or space scale (scaleful). Usually these modules are computationally intensive and would require HPC resources. They are often (but not always) implemented as parallel programs.
- software modules that prepare data for scaleful modules and process data from scaleful modules. Usually these modules do not have demanding computational requirements. However, to avoid communication overhead, they often require to be executed “close” to the scaleful modules they are connecting (they could even be implemented in the same process as the one for the scaleful modules).

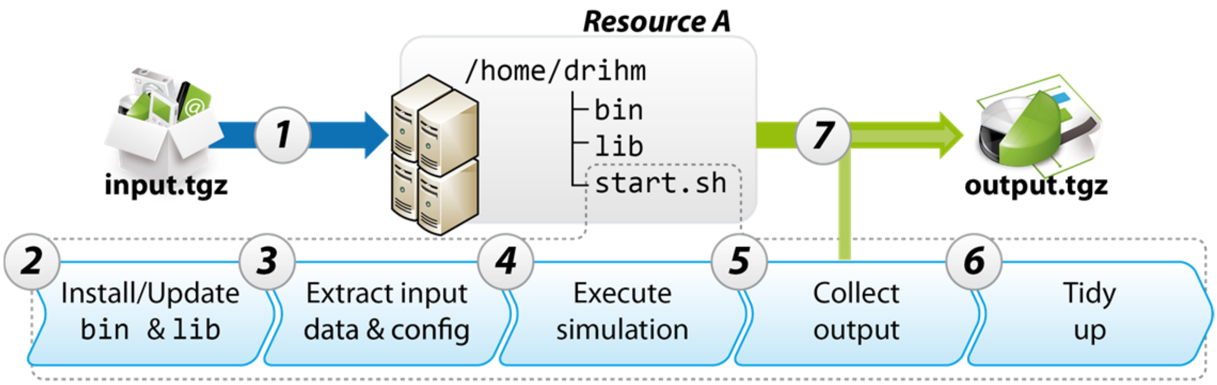

Figure 4: Typical HMR Application (DRIHM Example)

The communication structures of HMR applications may vary significantly between:

- a master-slave paradigm, where a master module triggers the execution of subordinate ones,
- a peer-to-peer type of computation where all modules are executed concurrently and exchange data in a usually asynchronous fashion. During the course of their execution, applications often pass many synchronization points (the number can be static or dynamic). Therefore, this type often requires mechanisms for efficient communication.
- a pipe type communication where modules execute one after another.
- a hybrid communication with combinations of two or more types mentioned above. The initial condition module is connected to the rest of the simulation via “pipe”, and then the rest of the simulation consists of modules that run concurrently.

Regarding the HMR Base Case several out-of-band/out-of-scope considerations apply, they need to be specified in detail. Examples are:

- resource descriptions
- fault tolerance model
- security mechanisms
- directory services
- systems management

- scheduling policies
- resource reservation

Given the HMR Base Case, the HMR Basic Profile describes the normative and non-normative requirements a service provider needs to meet. The sum of these requirements defines the HMR Basic Profile. In particular, the profile must specify

- Basic execution services
- Job descriptions
- Data staging requirements
- Security requirements

Claims of conformance to the HMR Basic Profile are made using the mechanisms described in the WS-I Conformance Claim Attachment Mechanisms [5].