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

D2.1: Report on an Assessment of Current e-Infrastructure Approaches for Hydro-Meteo Research in Europe and the US

Abstract: This document reports on an assessment of selected e-infrastructure approaches for hydro-meteo research in Europe, the US and Australia. The report is a deliverable of Task 2.1.

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	2.1
Deliverable Name	Report on Assessment of HMR e-Infrastructure Approaches
Work Package Number	2
Work Package Name	Architecture Harmonization Analysis and Planning
Deadline	01/08/2013
Version	1.0
Dissemination Level	PU
Nature	R
Lead Beneficiary	LMU

Document History

Date	Version	Description
15th Jul 2013	1.0	Initial write up for review by the project team.
29th Aug 2013	1.0	Integration of reviewer's comments

Table of Contents

List of Figures

1 Executive Summary

The purpose of work package 2 (Architecture Harmonization Analysis and Planning) is to analyse different e-infrastructure approaches in Europe and in the US as far as they relate to hydro-meteorological research. There are several ways to conduct such an analysis. This report is based on the proposal of a reference framework against which the e-infrastructures to be assessed are mapped. The framework considers different tiers of interest:

1. On a core infrastructure tier the building blocks consist of the resources, their deployment environments and the middleware stacks to access the resources. The stakeholders focusing on this tier are typically system-oriented ICT people.
2. On a Hydro-Meteorological Research (HMR) community tier the framework distinguishes between common services generic to *all* scientific communities and HMR specific services. While the former ones deal with model coupling mechanisms, workflow execution and data provenance, the latter ones support HMR specific data management, workflow generation and scientific annotations. Issues to solve on this tier belong to fields like interoperable geospatial cataloguing, data and model integration and service compatibility. This tier also deals with HMR models and applications. The stakeholders focusing on this tier are typically application-oriented HMR scientists and ICT stuff.
3. On a dedicated user tier the focus is on accessing the building blocks of the other tiers. Consequently, it considers the various client aspects like command line client, portals, graphical user interfaces, or application programming interfaces.

Given this framework, several approaches and contributions to e-infrastructures as far as they relate to HMR are analysed. In particular these are:

- The multiscale service stack provided by the EU-funded MAPPER project (Multiscale Applications on European e-Infrastructures)
- The research infrastructure provided by the EU-funded DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology)

- The earth system model provided by the US-funded CESM project (Community Earth System Model)
- The earth system model provided by the US-funded ESMF project (Earth System Modeling Framework)
- The hydrologic information system provided by the US-funded HIS project (Hydrologic Information System)
- The framework provided by the Australian WIRADA project (Water Information Research and Development Alliance)

The outcome of this task can be summarized as follows:

1. There is a growing landscape of services and tools, which individually are valuable for HMR scientists.
2. There is no "best HMR e-infrastructure" on the horizon, but best practices are.
3. A "Best Practices HMR e-infrastructure" can be pieced together even now.
4. The services and tools of HMR related e-infrastructures are a priori not interoperable.
5. There is a growing need for standardization but the standardization process is not fast enough.

The findings of task 2.1 will now be fed into task 2.2 (Derive a Generic Model) to define a generic architectural model, which will later serve as input to the gap analysis (task 2.3). It should be noted that the proposed framework canonically defines an architectural model where the gaps can easily be derived from the framework cells. For more information on this topic we refer to the respective reports D2.2 (Report on a common architecture model), D2.3 (Opportunity and gap analysis) and D2.4 (Future integration report).

2 Introduction

2.1 Background

The emergence of data driven science reflects the increasing value of observational, sensor, streaming and experimental data in every field of science. Hence, information and communication technology infrastructures supporting scientific data handling and workflow processes are gaining more and more importance. However, often data can neither be shared nor are data consuming applications and models interoperable across countries and disciplines; moreover, both are mostly unsustainable due to the lack of commonly agreed governance policies, inadequate legal frameworks and short-sighted funding models [7].

It is envisioned that by 2020/2030 all stakeholders, from scientists, science managers, infrastructure operators and governmental authorities to the general public, are aware of the critical importance of preserving and sharing reliable data collected by a vast array of sensors and instruments during scientific processes and everyday life behaviour [9] [14]. Unfortunately, however, diversity is likely to remain a dominant feature of scientific information: not only diversity of formats and types but also of people and communities that generate and use scientific data.

It is therefore of paramount interest that user communities in the various fields of science and humanities, research facilities producing data, and international teams of scientists should take a more active role in the definition of concrete short and long term requirements for the underlying scientific data infrastructures. Federated repositories, libraries and data centres should be interoperable at a global level with high degrees of dependability and trust, guided by international standards. Additionally, because all research builds on earlier work, not only fuller and wider access to scientific data is a prerequisite for accelerating scientific progress, but also a well-concerted interplay between all stakeholders as exemplified in Figure 1 for the European Grid Infrastructure (EGI) ecosystem [10].

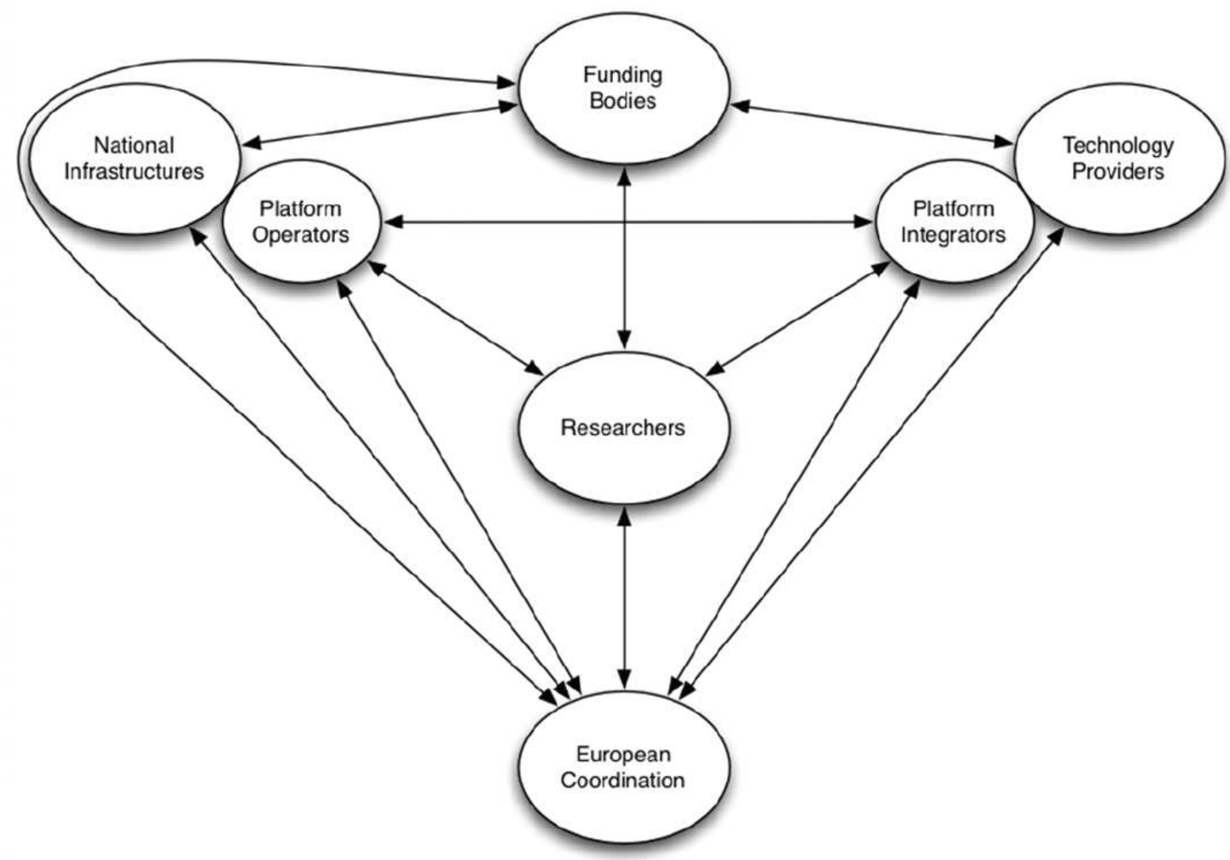


Figure 1: EGI Ecosystem [10]

The overarching goal of the DRIHM2US project is to support this vision. DRIHM2US aims at promoting the international cooperation between Europe and the USA for the development of a global, interoperable, sustainable common e-infrastructure for Hydro-Meteorological Research (HMR) by facilitating the persistent availability and an effective sharing of data and models across scientific disciplines, institutions and national boundaries [1].

HMR is an area of critical scientific importance and of high societal relevance. It plays a key role in guiding predictions relevant to the safety and prosperity of humans and ecosystems from highly urbanized areas to coastal zones and agricultural landscapes. Although it is well understood that hydro-meteorological events have to be investigated along the geographic

scale (local, regional, global) [2], it is necessary to better understand the commonalities and differences of e-infrastructures supporting this endeavour and to eventually define a "best practice" HMR e-infrastructure. Several projects and initiatives have recently been proposed or funded to support this vision. In this report we will discuss the following ones:

- The multiscale tools and services stack provided by the EU-funded MAPPER project (Multiscale Applications on European e-Infrastructures[1])
- The research infrastructure provided by the EU-funded DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology[2])
- The earth system model provided by the US-funded CESM project (Community Earth System Model[3])
- The earth system model provided by the US-funded ESMF project (Earth System Modeling Framework[4])
- The hydrologic information system provided by the US-funded HIS project (Hydrologic Information System[5])
- The framework provided by the Australian WIRADA initiative (Water Information Research and Development Alliance[6])

---

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

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

[3] http://www2.cesm.ucar.edu/

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

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

[6] http://www.csiro.au/en/Organisation-Structure/Flagships/Water-for-a-Healthy-Country-Flagship/WIRADA_WFHC_ResearchProfile.aspx

This report is organized as follows: We assess the various projects and initiatives separately to gain a better understanding of their peculiarities in section 3. For the assessment itself we first define relevant terms before outlining a reference framework to be used for the assessment). This section also contains a brief ratio as far as the e-infrastructure candidate selection is concerned. Section 4 discusses the findings of the assessment exercise and section 5 concludes the report.

2.2 Remarks

This report is written with some constraints in mind:

- While section 3 is kept short for better readability, the majority of the assessment material is collected in the appendix.
- For the various project descriptions we refer to material available from the respective web sites without explicit quoting the passages. The source of the quotes, however, should be obvious from the context.

3 Assessment

The assessment presented in this report is based on three major sources: a literature screening, an analysis of respective project deliverables, and the output presented at the DRIHM2US expert networking sessions of work package 3. The latter is reported in deliverable D3.3 (Domain expert networking report) [3].

3.1 Terms and Methodology

Research in general and HMR in particular is increasingly based on distributed regional, national and global collaborations of scientists facilitated by the Internet as a global communication infrastructure and by cooperation mechanisms following the Grid Computing paradigm [4], the Cloud Computing paradigm [5] or a mix of both.

"e-Infrastructure" is the term used for the technology and the organisational model that support research undertaken in this way [6]. Consequently, it embraces networks, Grids, data centres, collaborative environments, service registries, single-sign on mechanisms, certificate authorities, training and help-desk services. Most importantly, however, it is the seamless integration of these concepts that defines an e-infrastructure. Community-specific e-infrastructures (like the ones relevant for HMR) also feature community-specific mechanisms like model coupling or data format conversions.

In its broadest possible terms, e-infrastructures thus deal with different stakeholders, data types and services that somehow interrelate to facilitate global science processes. Data generators (like sensors or instruments) and users (like scientists or citizen scientists [14]) gather, capture, transfer and process data - often across the globe in virtual research communities. They draw upon support services in their specific scientific communities (e.g., HMR), which typically comprise tools to locate data, process it, annotate it or interpret it. The support services themselves make use of a broad set of common data services including data storage, data identification, data authentication, data mining, and workflow/task execution.

Based on these observations we propose the generic reference framework of Figure 2 which also reflects the findings and suggestions in

- The European Commission High Level Expert Group on Scientific Data report "Riding the Wave" from October 2010 [13]
- The European Commission report "Open Infrastructures for Open Science - Horizon 2020 consultation report" by Richard L. Hudson and Carlos Morais Pires [9]
- The 2009 White Paper "Strategy for a European Data Infrastructure" of the European data initiative PARADE (the Partnership for Accessing Data in Europe) [15]
- Wilkins-Diehr, Nancy, Dennis Gannon, Gerhard Klimeck, Scott Oster and Sudhakar Pamidighantam: TeraGrid science gateways and their impact on science, published in Computer, 41(11):32{41, 2008} [11]
- The XSEDE architecture ratio as described in [16]
- The Second European Union-Australia Workshop on Research Infrastructure [8]

The framework follows a simple 3-tier approach:

1. The User Tier comprises all building blocks relevant to the user working with the HMR e-infrastructure. Note that users may be scientists, administrators, citizen scientists, software processes, and sensors. Accordingly, access may be facilitated by command line interfaces, portals, GUIs, APIs, or alike.

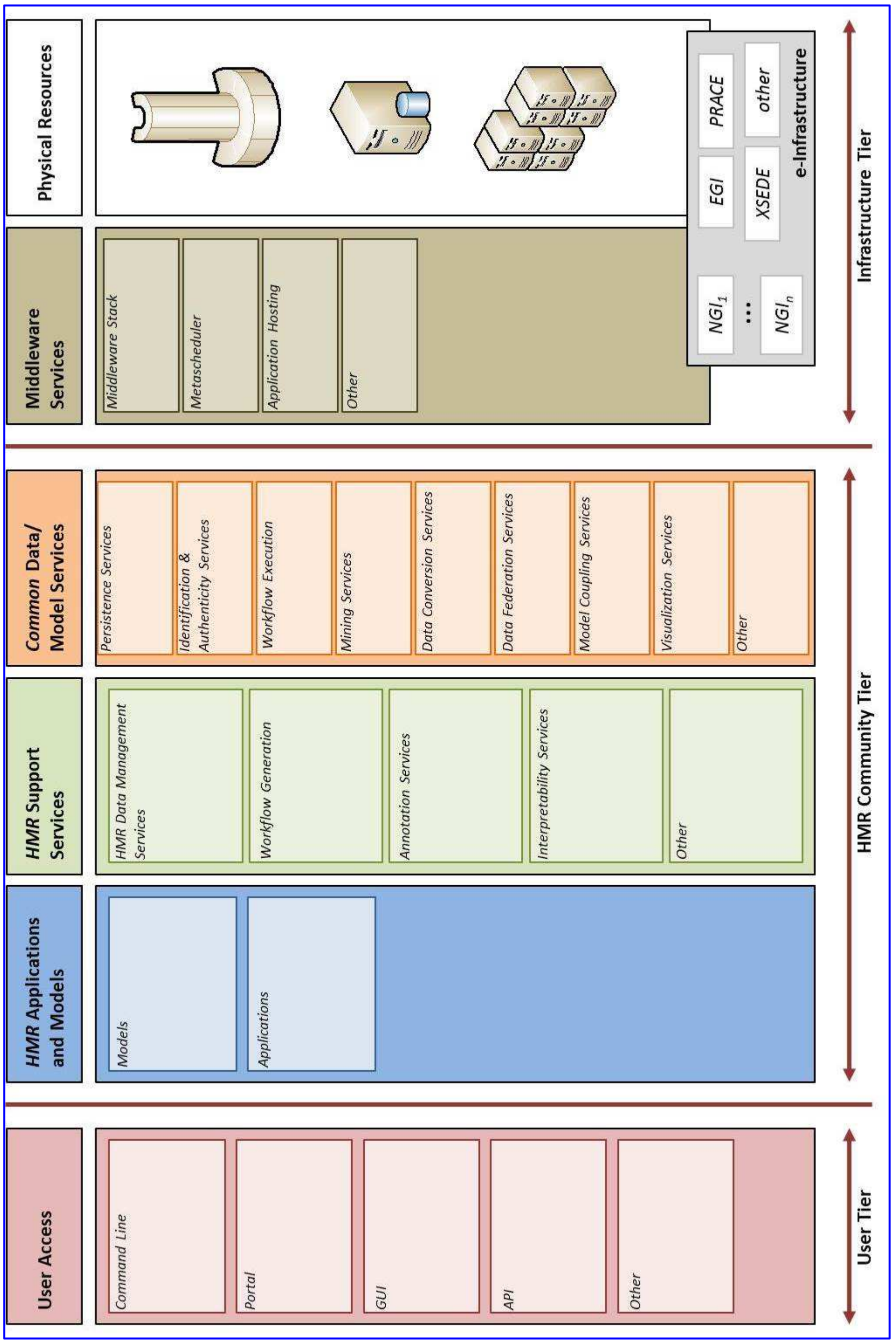


Figure 2: Generic Reference Framework

2. Applications somehow use models and data. This usage pattern is reflected in the HMR Community Tier building blocks where applications draw upon HMR specific support services. The support services themselves make use of common (i.e., generic) data/model services that include functionalities to store and identify data (including archiving and Persistent Identifiers[7]), authenticate it, execute tasks, and mine it. Common services also include data conversion tools, model coupling mechanisms and generic visualization methods.
3. The Infrastructure Tier provides the physical resources and the mechanisms to access and manage these resources. The resources may in total or in parts be "contributed to" the National Grid Infrastructures (NGI), to the European Grid Infrastructure (EGI), to the Partnership for Advanced Computing in Europe (PRACE), to the Extreme Science and Engineering Discovery Environment (XSEDE) or others. Access is typically enabled by middleware technology assisted by metascheduler technology and application hosting environments -- and preferably all based on standards.

This report cannot assess *all* e-infrastructures and their service stacks that may relate to HMR. Instead, we propose a classification of infrastructure approaches along very basic criteria (see Figure 3). E-infrastructures[8] fall into two main classes:

1. Basic provisioning and management infrastructures with core communication networks, computing and storage facilities, and services necessary for accessing these resources;
2. Infrastructures consisting of toolboxes for data/model management.

Typically, the latter ones augment the former ones. Basic infrastructures are often realized over Grids, Clouds, or Web Services. Following the generic framework (Figure 2), the science

---

[7] e.g., http://www.handle.net/

[8] as far as they are of importance for this report

enabling infrastructures can again subdivided into generic toolboxes (for data management and model management) and science (here HMR) specific toolboxes.

In this report we will not discuss the category of basic infrastructures in detail. They are well-known and many details may be found in the literature. In particular, we refer to http://www.egi.eu/ for information about EGI, to http://www.prace-ri.eu for details regarding the PRACE Research Infrastructures, and to https://www.xsede.org/ for information regarding the US XSEDE initiative.

Please note also that the scheme in Figure 2 does not include any application-based classification. Instead, applications can be subsumed under the science specific toolboxes. Thus, given the classification scheme of Figure 3 we look into

- The MAPPER project as it provides generic tools.
- The DRIHM project as it provides both generic and specific tools across multiple science disciplines.
- The CESM project as it provides a fully-coupled, global climate model with state-of-the-art simulations.
- The ESMF project as it provides generic tools for building climate, numerical weather prediction, data assimilation, and other Earth science applications.
- The CUAHSI HIS project as it provides tools for sharing hydrologic data.
- The Australian WIRADA initiative as it provides both common and water specific tools based on standards.

The assessment criteria are obvious from the "service boxes" in Figure 2and Figure 3.

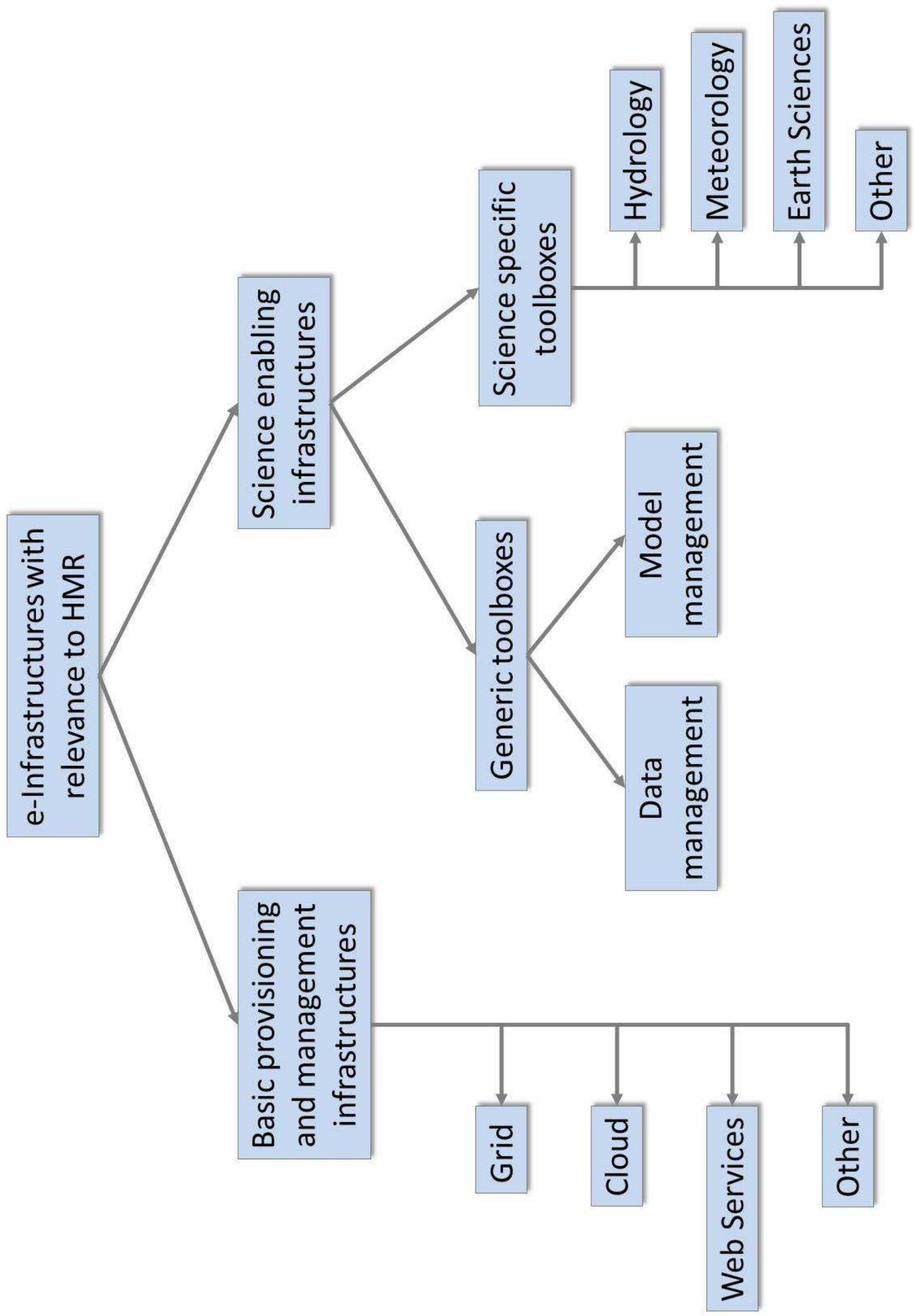


Figure 3: Classification scheme for e-Infrastructures

3.2 Assessment of HMR Related e-Infrastructures

In this section we briefly analyse selected e-infrastructures. For every candidate we first give a short description of the project before mapping the project onto the generic reference framework of Figure 2. Note that this mapping is performed to the best of our knowledge based on information available as of writing. Should there emerge additional or better information later, this will be reflected in an addendum to this report and in the subsequent reports D2.2, D2.3, and D2.4. Note also the remarks in section 2.2.

3.2.1 Distributed Research Infrastructure for Hydro-Meteorologic Research (DRIHM)

Description

Hydro-meteorological science has made strong progress over the last decade at European and worldwide levels: new models, enhanced post-processing methodologies and observational datasets are available to tackle hydro-meteorological research questions. A fundamental challenge is the dynamic coupling and integration of dedicated Grid and HPC services to facilitate a multi-disciplinary and global collaboration between meteorologists, hydrologists and other Earth science experts.

The DRIHM (Distributed Research Infrastructure for Hydro-Meteorology) project aims at deploying such an e-infrastructure by provisioning end-to-end HMR services (models, datasets and post-processing tools) at European and possibly global levels. The objectives are to help the definition of a common long-term strategy, to foster the development of new HMR models and observational archives for the study of severe hydro-meteorological events, to promote the execution and analysis of high-end simulations and analysis, and to support the dissemination of DRIHM results. The forecasting chains are reflected in Figure 4[9].

[9] Courtesy Quillon Harpham

DRIHM combines the European expertise in HMR modelling, in hydro-meteorological remote sensing, in Grid and High Performance Computing (HPC) science, and in studies of severe weather impacts.

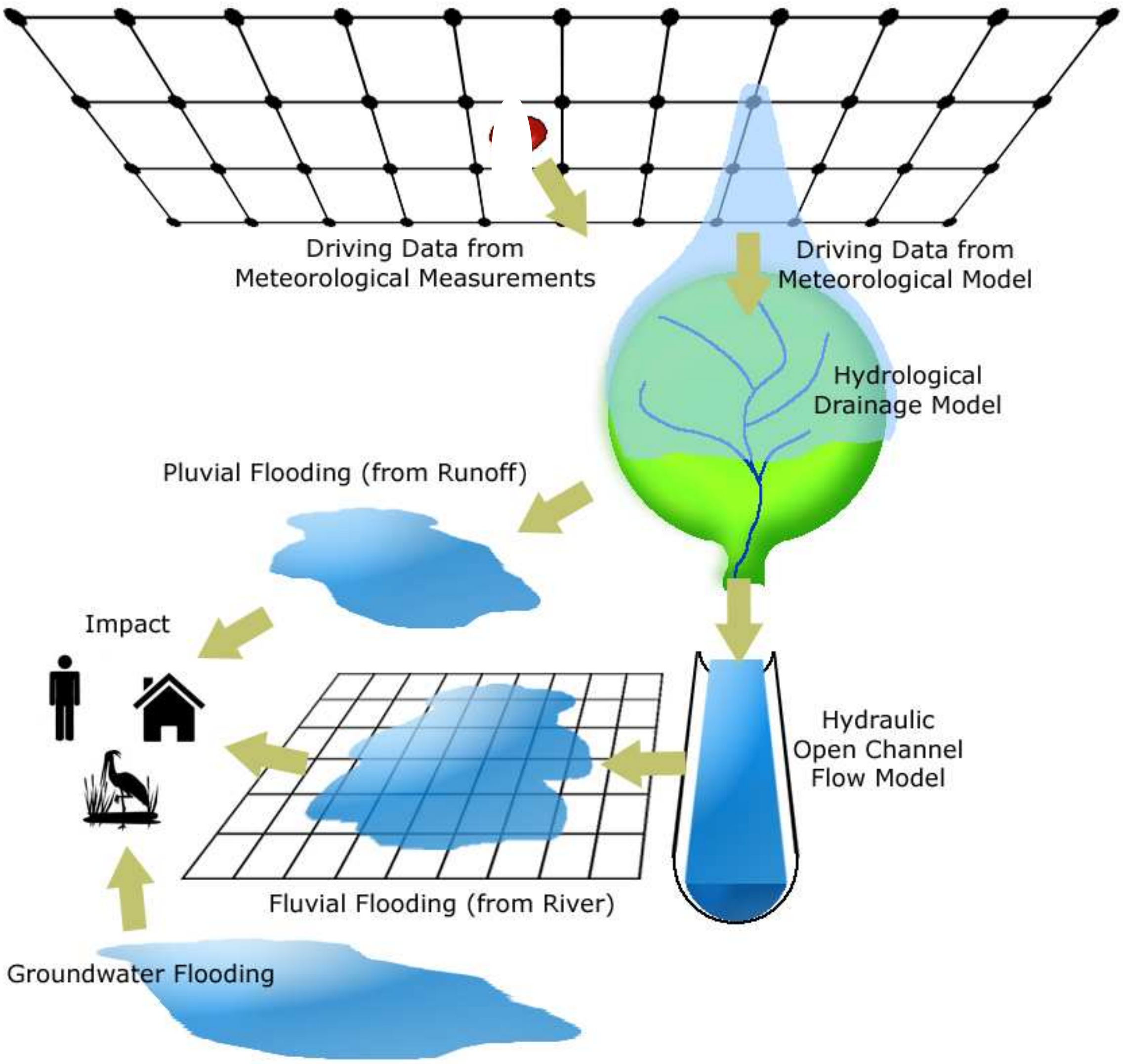


Figure 4: DRIHM Forecasting Chains

Framework Mapping

The DRIHM framework mapping is summarized in Appendix 1. For more information we refer to the project's home page http://www.drihm.eu/.

DRIHM services are primarily deployed on Grid infrastructures as provided by NGIs and EGI and a roadmap towards PRACE. Portal driven, DRIHM integrates/couples several models to run scientific HMR workflows. Model coupling is mainly facilitated by scripting and standards-based data management (e.g., WaterML). The DRIHM concept is flexible enough to not only support extensions of the HMR model landscape but also to support adaptations to changes in the infrastructure.

As of this writing, a complete execution of the full forecasting chain has not been executed in production.

The current status of the implementation of the DRIHM infrastructure is summarized in Figure 5 with the (preliminary) repository, the portal, the Grid interfaces, and the EGI/PRACE embedding. As of this writing, major building blocks of the reference framework are available. While science specific services (e.g., mining services, annotation services, interpretability services) are not in the focus of the project, they are "borrowed" from other projects.

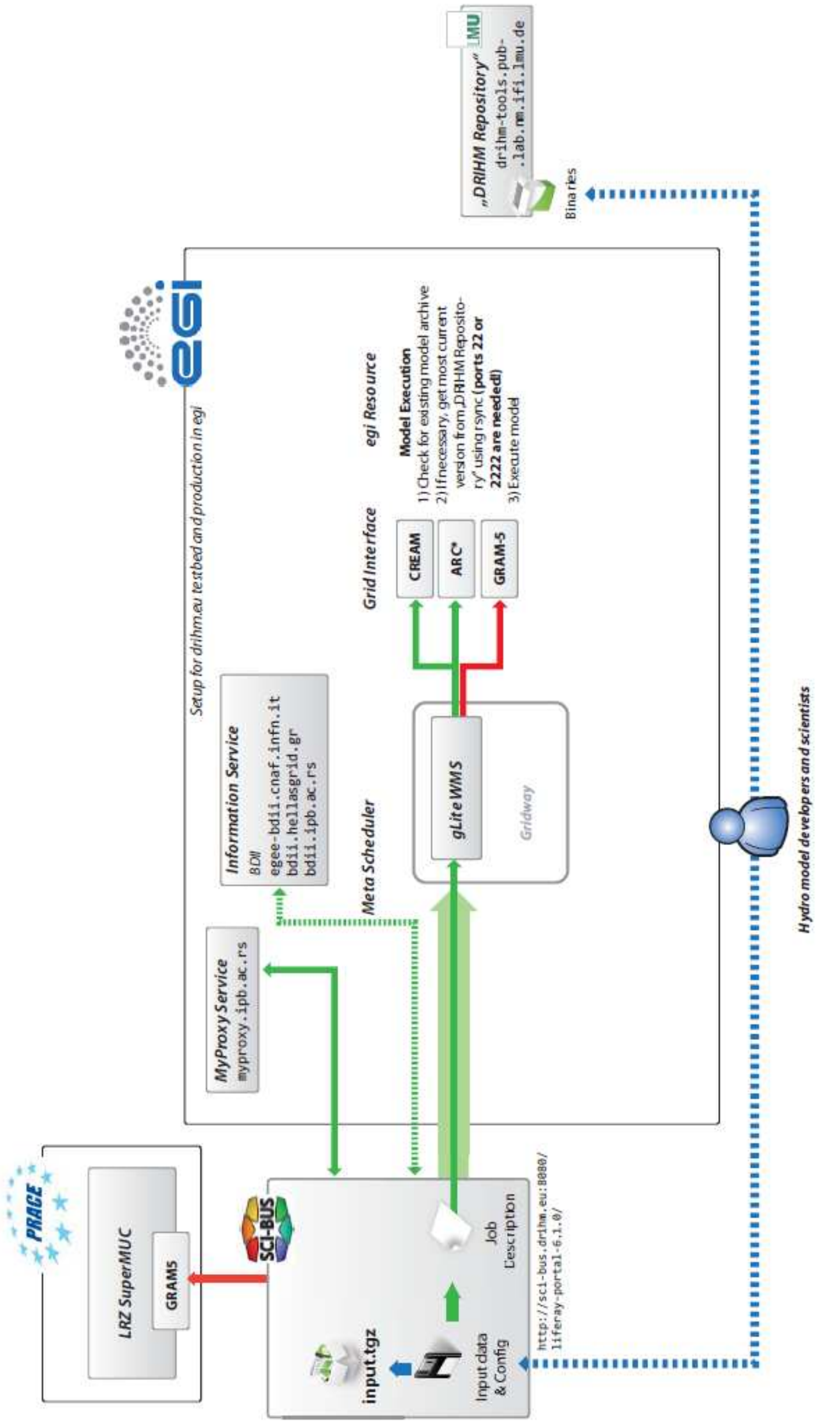


Figure 5: DRIHM Infrastructure Status (as per end August 2013)

3.2.2 Community Earth System Model (CESM)

Description

The Community Earth System Model (CESM) is a coupled climate model for simulating the Earth's climate system. Composed of separate models simultaneously (atmosphere, ocean, land, land-ice, sea-ice) plus one central coupler component, CESM allows researchers to conduct fundamental research into the Earth's past, present, and future climate states. CESM is funded by the US National Science Foundation (NSF) and the U.S. Department of Energy (DOE). Administration of the CESM is maintained by the Climate and Global Dynamics Division (CGD) at the National Center for Atmospheric Research (NCAR).

CESM addresses important areas of climate system research by aiming at better understanding and predicting climate phenomena. The long-term goals of the CESM project are simple but ambitious (see also http://www2.cesm.ucar.edu/about). They are:

- to develop and to work continually to improve a comprehensive earth system model that is at the forefront of international efforts in modelling the climate system, including the best possible component models coupled together in a balanced, harmonious modelling framework;
- to make the model readily available to, and usable by, the climate research community, and to actively engage the community in the on-going process of model development;
- to use the model to address important scientific questions about the climate system;
- to use appropriate versions of the model for calculations in support of national and international policy decisions.

Framework Mapping

The CESM mapping is summarized in Appendix 2. For more information we refer to the project’s home page http://www2.cesm.ucar.edu/.

CESM provides scientific validations that consist of multi-decadal model runs of a given

component set at a given target resolution, followed by a scientific review of the model output diagnostics. All scientifically validated component sets are accompanied by diagnostic and model output data. The validated configurations are available from https://www2.cesm.ucar.edu/models/scientifically-supported.

Because CESM aims at focusing on the Earth System Model in order to better understand the Earth's dynamics over long time periods, there is an "eco system" of supporting services to make these simulation runs happen. Although CESM is not devoted to HMR, HMR simulation may however benefit from the data/models available. In particular, CESM's standards based approach may simplify the integration of CESM data/models into HMR applications and workflows.

3.2.3 Earth System Modeling Framework (ESMF)

Description

Similar to CESM, the Earth System Modeling Framework (ESMF) collaboration is aiming at building a high-performance, flexible software infrastructure to improve climate research, numerical weather prediction, data assimilation, and corresponding Earth science applications. ESMF defines an architecture for composing complex, coupled modelling systems and includes data structures and utilities for developing individual models (see Figure 6, borrowed from http://www.earthsystemmodeling.org/about_us/).

The basic idea behind ESMF is that complicated applications should be broken up into smaller pieces, or components. A component is a unit of software composition that has a coherent function, and that exposes a standard interface and behaviour. Components can be assembled to create multiple applications, and different implementations of a component may be available.

ESMF also includes toolkits for building components and applications, such as regridding software, calendar management, logging and error handling, and parallel communications.

For more information on ESMF we refer to http://www.earthsystemmodeling.org/index.shtml.

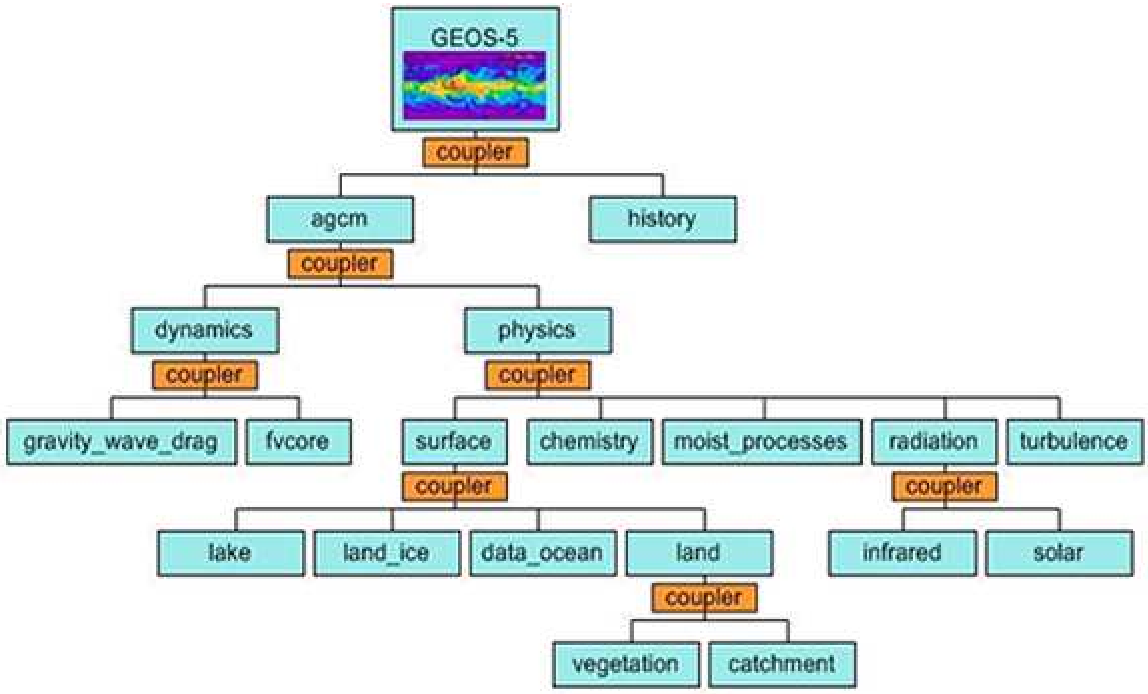


Figure 6: ESMF General Architecture (Example)

Framework Mapping

The ESMF mapping is summarized in Appendix 3. For more information we refer to the project's home page http://www.earthsystemmodeling.org.

ESMF provides a so called superstructure for assembling geophysical components into applications and an infrastructure that modellers use to share common utility code for functions like grid remapping and time management. Via the National Unified Operational Prediction Capability (NUOPC) layer a set of calls exists to aggregate ESMF functions and to encode metadata, sequencing, and other conventions to improve interoperability and make ESMF workflows simpler (see Figure 7).

ESMF provides with the Common Information Model (CIM) to describe climate data and the models that produce it in a standard way (NetCDF CF). Additionally, in two-way coupling, each ESMF model pulls data from other models using OpenMI.

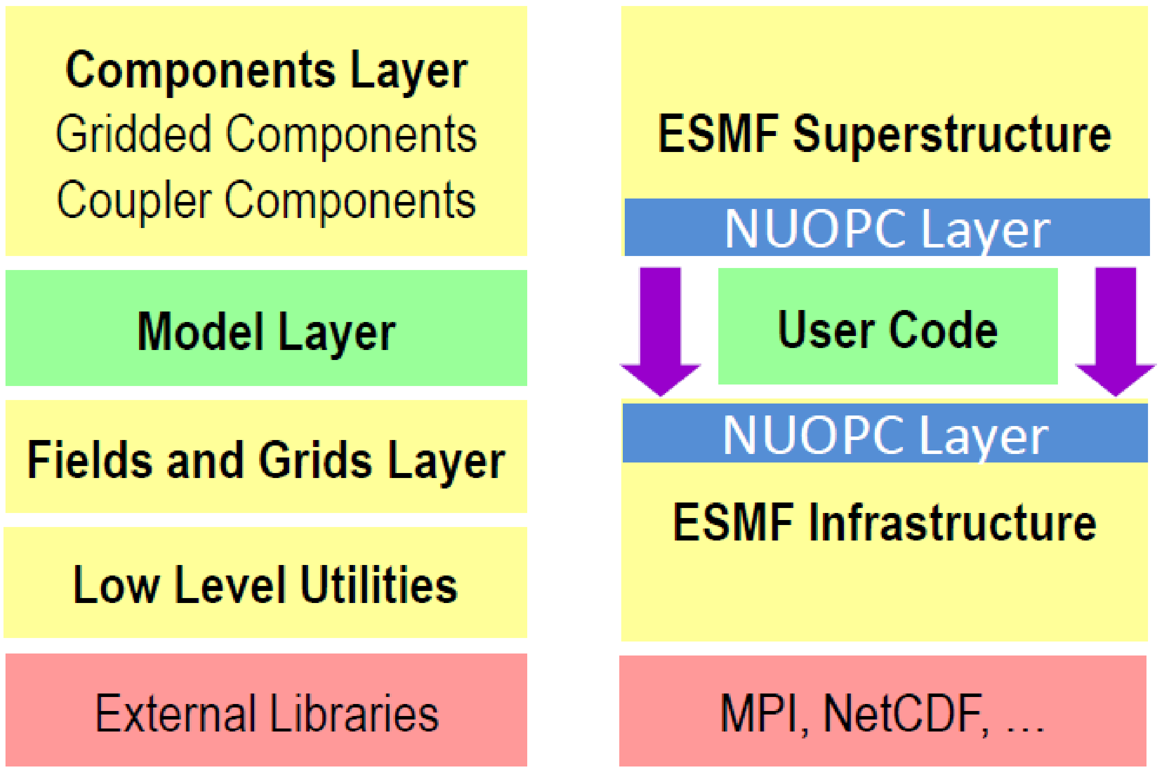


Figure 7: ESMF Structure

As CESM, ESMF provides a modelling framework, however, with the focus on geophysical phenomena rather than HMR ones.

3.2.4 Hydrologic Information System (HIS)

Description

The CUAHSI Hydrologic Information System (HIS) is an internet-based system for sharing hydrologic data. It is comprised of databases and servers, connected through Web Services to client applications, allowing for the publication, discovery and access of data. [12]

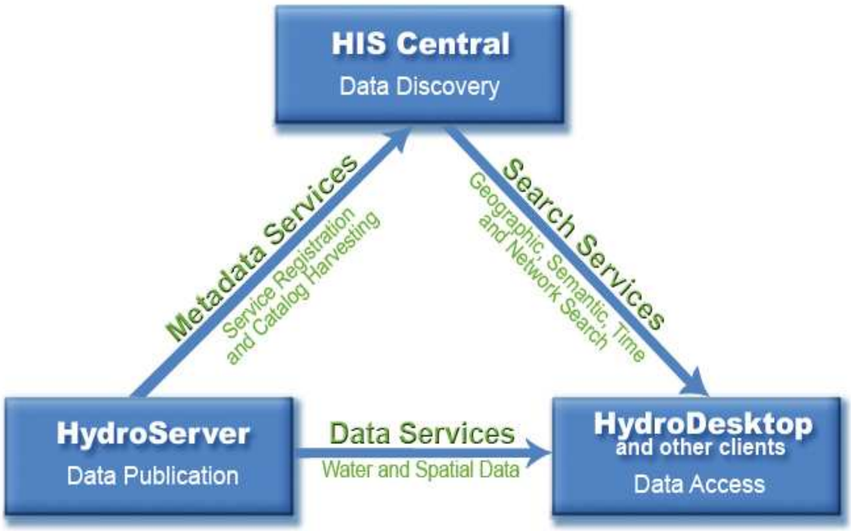


Figure 8: HIS Components

There are three types of systems that store and process data in HIS (see Figure 8):

- HIS Central contains copies of metadata, which facilitates searches. It works like a search engine, in that it harvests metadata from the data servers and allows it to be efficiently searched by the clients.
- HydroServer stores, organizes and publishes data. It allows metadata to be harvested by HIS Central and data to be shared with clients.
- Clients (such as HydroDesktop) give users a convenient interface to access data. They retrieve metadata from HIS Central and retrieve data from HydroServers.

HIS provides three types of web-services that allow for the communication via the internet:

- Data Services allow water and related spatial data to be retrieved by the client (such as HydroDesktop) computers.
- Search Services allow the client computers to perform searches of the search catalogue at HIS Central.
- Metadata Services allow HIS Central to retrieve the metadata necessary to build the search catalogue.

For more information on CUAHSI HIS we refer to http://his.cuahsi.org/.

Framework Mapping

The HIS mapping is summarized in Appendix 4. For more information we refer to the project's home page http://his.cuahsi.org/.

HIS is "operated" through graphical user interfaces (HydroDesktop, HydroExcel) and Web Services in the context of pure hydrologic science. Accordingly, HIS provides standardized ways to convert scientific data between models (WaterML), to couple models (OpenMI), to retrieve and integrate data using standard templates (Observations Data Model (ODM)).

In the centre of the CUAHSI HIS architecture is HIS Central, a web application that provides interfaces for adding and managing registered water data services and the HIS Central Metadata Catalog. The catalogue is designed to maintain observation series information, including site information, variable information, the period of record, as well as project metadata – for all registered data sources of hydrologic observations.

3.2.5 Multiscale Applications on European e-Infrastructures (MAPPER)

Description

Driven by seven challenging applications from five representative scientific domains (fusion,

clinical decision making, systems biology, nano science, engineering), MAPPER deploys a computational science environment for distributed multiscale computing on and across European e-infrastructures. By taking advantage of existing software and services, as delivered by European and national projects, MAPPER develops tools, software and services that permit loosely and tightly coupled multiscale computing in a user friendly and transparent way.

MAPPER integrates heterogeneous infrastructures for programming and execution of multiscale simulations. MAPPER solutions are developed on top of existing e-infrastructures without the necessity to modify already deployed components. The functionality to be delivered is realized as extensions to existing e-infrastructures. The integration is done using well defined APIs and standard based interfaces, thus reducing potential impact of changes on middleware level components.

For more information on the MAPPER project we refer to http://www.mapper-project.eu.

Framework Mapping

The MAPPER mapping is summarized in Appendix 5. For more information we refer to the project's home page http://www.mapper-project.eu.

MAPPER is not HMR specific. Instead, MAPPER provides a set of generic tools and methods for distributed multiscale computing which may beneficially be applied to HMR specific applications as well because HMR problems inherently require a multiscale modelling approach. As multiscale models can be very computationally expensive, a distributed infrastructure may be beneficial. Therefore MAPPER provides with tools like the Submodel Execution Loop (SEL), standardized coupling templates, a well-defined Multiscale Modeling Language (MML), the Multiscale Coupling Library and Environment (MUSCLE) is outlined, the necessary support to run multiscale applications on distributed infrastructures like NGI, EGI and PRACE. [19]

As of this writing MAPPER is in its final project stage. The MAPPER sustainability plan is available at http://www.mapper-

project.eu/documents/10155/23400/D3.6_FinalSustainabilityPlan_v1_FINAL.pdf.

3.2.6 Water Information Research and Development Alliance (WIRADA)

Description

In 2008 the Australian Bureau of Meteorology and Australia's national science agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), joined their research forces to establish the Water Information Research and Development Alliance (WIRADA). Since then WIRADA has delivered significant achievements in improving Australia's water information systems, water accounting, and water forecasting, with national and global impacts.

WIRADA developed a reliable seasonal streamflow forecasting service (see Figure 9 and [17]), a high-resolution digital elevation model for landscape and water resources, a national water resources modelling system, and a common format within an international water data standard.

Framework Mapping

The WIRADA mapping is summarized in Appendix 6. For more information we refer to the project's home page http://www.csiro.au/Organisation-Structure/Flagships/Water-for-a-Healthy-Country-Flagship/WIRADA_WFHC_ResearchProfile.aspx.

The primary access to all WIRADA services is the Hydrologists Workbench facilitating an automated management of the flows of exponentially increasing data volumes through models, while ensuring auditability and compliance with standards (WaterML, Water Data Transfer). Built on commercial off-the-shelf scientific workflow software[10] which provides the workflow, audit and governance utility, the Workbench draws together public domain and proprietary

10 Kepler (https://kepler-project.org/) was trialled, Trident (http://tridentworkflow.codeplex.com/) was selected. See [27].

hydrological, statistical and geo-information system toolkits with tailored workflows to provide an extensible portal for the provision and management of (one-off or routine) modelling exercises.

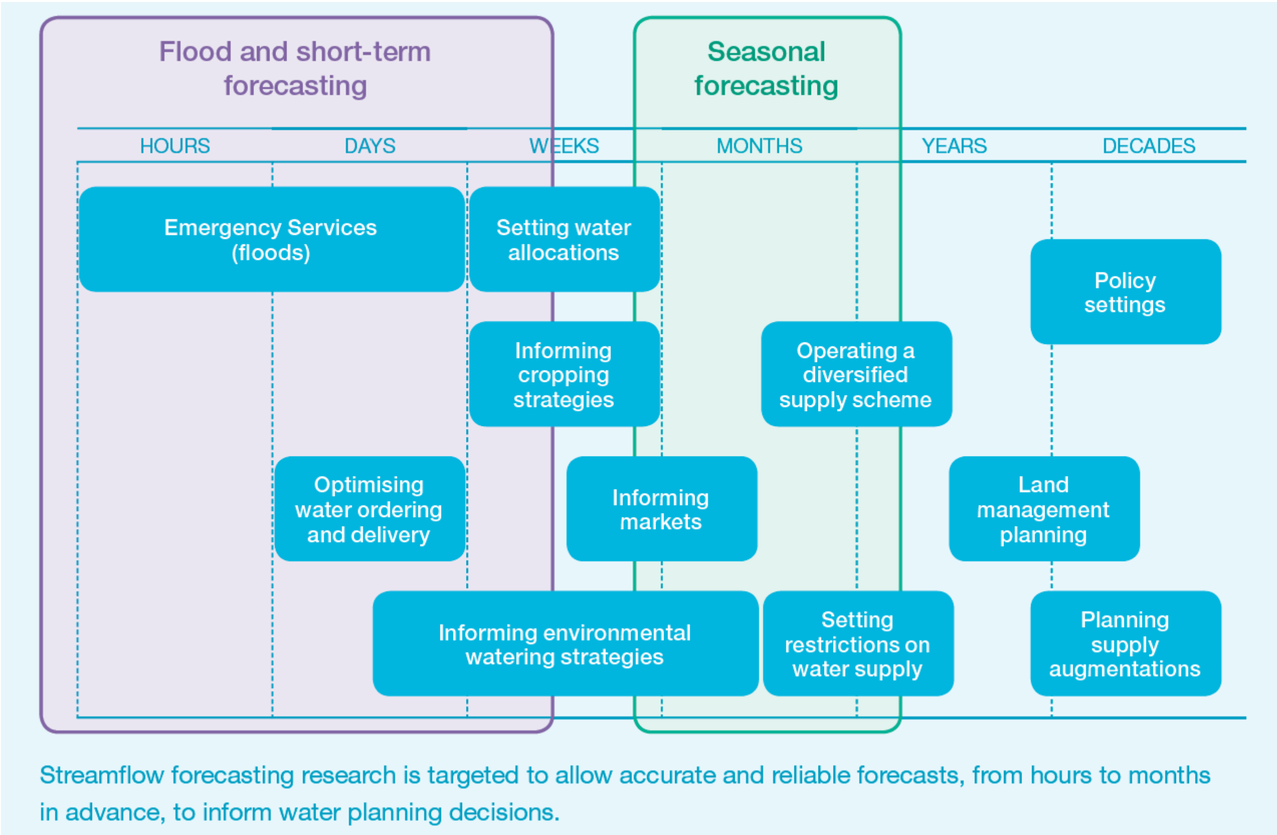


Figure 9: WIRADA Streamflow Forecasting [17]

Figure 10 (from [20]) summarizes the functionality provided by the Workbench. Note the HIS adapter, the Australian Water Data Infrastructure Project (AWDIP)[11] to develop interoperability

[11] http://www.daff.gov.au/brs/water-sciences/ground-surface/awdi-project

standards and protocols, the IQQM river system model[12], and the TIME software development framework for creating models[13].

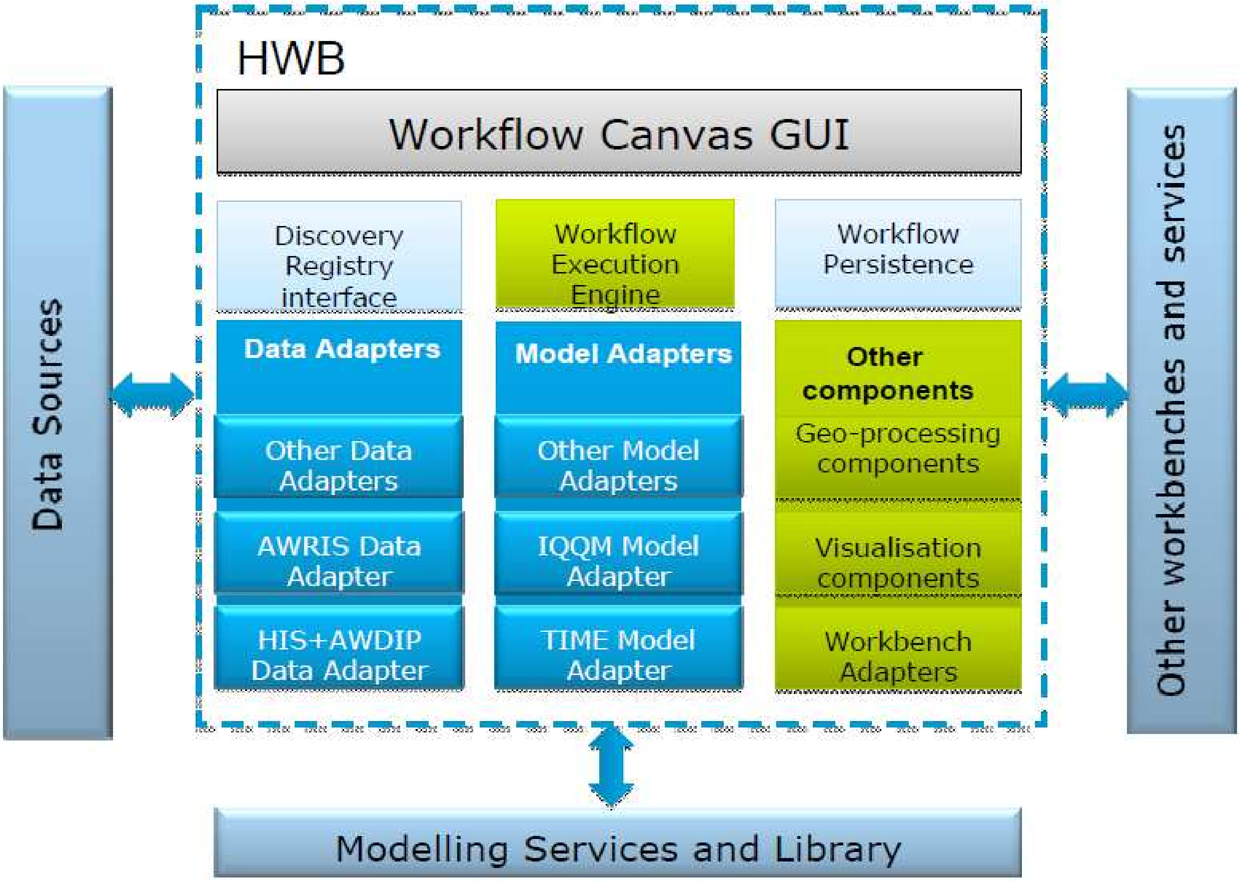


Figure 10: Functionality Provided by Hydrologists Workbench [20]

---

[12] http://www.sciencedirect.com/science/article/pii/S0266983896000196#

[13] http://www.toolkit.net.au/Tools/PublicationDetail.aspx?id=1000015&publicationID=1000016

4 Findings Summary

Section 3 and the related appendices contain a rough assessment of e-infrastructures (or e-infrastructure components) as far as they related to HMR. The outcome of the assessment can be summarized as follows referring to Figure 2:

1. The basic infrastructures on the Infrastructure Tier (NGI, EGI, PRACE, XSEDE) are assumed to be operational or there is no dependency on such infrastructures at all.
2. On the HMR Community Tier several models and applications are available. There is a large variety of HMR Support Services available, mostly "case-specific" but probably expandable. HMR-specific annotation services and interpretability services are not available, only rudimentary. Although common data/model services are available in various constellations, they are generally not generic and interoperable enough. Some projects offer very specific tools (e.g., CESM) while others try to support a broader spectrum (e.g., MAPPER).
3. The User Tier is typically project/initiative related because it facilitates the access to the HMR Community Tier. Nonetheless, it would be nice to have a generic HMR user API or a common HMR portal system for all HMR scientific activities over e-infrastructures.

The following table briefly summarizes the previous assessment. Please note the addition of the "Standards Based" category and the focus. The legend of the cells is as follows:

--: not available
-: very rudimentary
0: available as part of the tools or services
+: available in large parts
++: full availability
*: several
n/a: not applicable
X: main focus

		Project/Initiative					
Tier	**Category**	**DRIHM**	**CESM**	**ESMF**	**HIS**	**MAPPER**	**WIRADA**
User Access	CLI	0	0	++	0	0	0
	Portal	++	+	+	0	0	0
	GUI	+	-	0	+	+	++
	API	0	+	+	+	+	0
HMR Applications and Models	Models	*	*	*	0	n/a	*
	Applications	*	*	*	0	n/a	*
HMR Support Services	Data Mgmt. Services	0	+	-	+	0	+
	Workflow Generation Services	+	0	+	--	0	+
	Annotation Services	0	-	-	--	--	-
	Interpreta-bility Services	-	--	--	--	--	--
Common Data/Model Services	Persistence Services	0	+	-	0	--	-
	Identification and Authenticity	-	0	0	0	0	--
	Workflow Execution	+	--	+	--	0	+
	Mining Services	--	+	--	0	--	--

		Project/Initiative					
Tier	**Category**	**DRIHM**	**CESM**	**ESMF**	**HIS**	**MAPPER**	**WIRADA**
	Data Conversion Services	+	0	+	0	0	-
	Data Federation Services	0	0	0	0	0	-
	Model Coupling Services	0	0	0	0	++	+
	Visualization Services	0	0	+	--	0	--
	Standards Based	0	-	++	+	0	++
Middleware Services	Grid Middleware	*	n/a	n/a	n/a	*	n/a
	Metascheduler	*	n/a	n/a	n/a	*	n/a
	Application Hosting	*	n/a	n/a	n/a	*	n/a
Physical Resources		*	*	*	*	*	*
e-Infrastructure		*	*	*	n/a	*	n/a
Community Focus	HMR	X					
	Hydrology				X		X
	Geophysics						
	General Earth Sciences		X	X			
	Generic					X	

5 Conclusion

In this deliverable we reported on an assessment of HMR related e-infrastructures performed as part of task 2.1. The assessment was based on the definition of a reference framework and the selection of reference candidates according to a coarse-grained taxonomy. The reference framework itself has been derived from related efforts of the European Commission High Level Expert Group, the European Commission Horizon 2020 efforts, the European data initiative PARADE, the TeraGrid science gateway investigations, the XSEDE architecture ratio description, and the European Union-Australia Workshop on Research Infrastructures.

The general one-sentence finding of the assessment is: There are HMR related e-infrastructure building blocks available, but they must be combined successfully for future HMR services and applications. Which building blocks to consider as candidates for future HMR activities and which gaps to close will be the focus of subsequent reports D2.2 (Report on a common architecture model), D2.3 (Opportunity and gap analysis) and D2.4 (Future integration report).

6 Acronyms and References

6.1 Acronyms and Abbreviations

Acronym / Abbreviation	Definition
API	Application Programming Interface
CAVE	Cave Automatic Virtual Environment
Cb-TRAM	**C**umulonim**b**us **Tr**acking **a**nd **M**onitoring
CESM	Community Earth System Model
CGD	Climate and Global Dynamics Division
CIM	Common Information Model
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science, Inc.
DOE	Department of Energy
DRIHM	Distributed Research Infrastructure for Hydro-Meteorology
DRIHM2US	Distributed Research Infrastructure for Hydro-Meteorology to United State of America
EGCF	European Globus Community Forum
EGI	European Grid Infrastructure
ESMF	Earth System Modeling Framework
GUI	Graphical User Interface
gUSE	Grid and Cloud User Support Environment
HBV	Hydrologiska Byråns Vattenbalansavdelning
HIS	Hydrologic Information System
HMR	Hydro-Meteorological Research
HPC	High Performance Computing

ICT	Information and Communications Technology
MAPPER	Multiscale Applications on European e-Infrastructures
Meso-NH	Mesoscale Non-Hydrostatic Model
MML	Multiscale Modeling Language
MUSCLE	Multiscale Coupling Library and Environment
NCAR	National Center for Atmospheric Reseach
NCAR	National Center for Atmospheric Research
NetCDF	Network Common Data Form
NetCDF CF	NetCDF Climate and Forecast
NGI	National Grid Initiative
NSF	National Science Foundation
NUOPC	National Unified Operational Prediction Capability
ODM	Observations Data Model
OGC	Open Geospatial Consortium
OpenMI	Open Modelling Interface
PARADE	Partnership for Accessing Data in Europe
PRACE	Partnership for Advanced Computing in Europe
SCI-BUS	Scientific Gateway Based User Support
SEL	Submodel Execution Loop
WaterML	Water Markup Language
WIRADA	Water Information Research and Development Alliance
WRF	Weather Research and Forecasting
WRF-NMM	Weather Research and Forecasting Nonhydrostatic Mesoscale Model
XSEDE	Extreme Science and Engineering Discovery Environment

6.2 References

[1] Parodi, A. et al.: DRIHM Description of Work (DoW), 2011.

[2] United Nations Office for Disaster Risk Reduction (UNISDR): From Shared Risk to Shared Value –The Business Case for Disaster Risk Reduction. Global Assessment Report on Disaster Risk Reduction. Geneva, Switzerland, 2013

[3] DRIHM2US Project: Domain Expert Networking Report. Deliverable D3.3, July 2013

[4] 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

[5] Wolfgang Gentzsch ; Burak Yenier: The UberCloud HPC Experiment: Compendium of Case Studies. http://www.hpcwire.com/whitepapers/2013-06-25/the_ubercloud_hpc_experiment:_compendium_of_case_studies.html

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

[7] Carnegie Group of G8 + 05 Science Advisers: G8+O5 GLOBAL RESEARCH INFRASTRUCTURE, SUB GROUP ON DATA, Draft Report, 2011, *http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/g8.pdf*

[8] European Commission: The Second European Union-Australia Workshop on Research Infrastructure, Brussels, 2012; *http://ec.europa.eu/research/infrastructures/pdf/2nd-aus-eu-ri-workshopreport-oct.pdf#view=fit&pagemode=none*

[9] European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social committee and the Committee of the Regions; Horizon 2020 - The Framework Programme for Research and Innovation, Brussels, 2011

[10] Sergio Andreozzi: EGI Feedback on D3 -Final Study Report (draft) "Developing impact measures for e-Infrastructures". RI-Impact -Dissemination Workshop, 2012, http://www.ri-impact.eu/ri-impact-wAssets/docs/EGI-Perspective.pdf

[11] Wilkins-Diehr, Nancy, Dennis Gannon, Gerhard Klimeck, Scott Oster and Sudhakar Pamidighantam: TeraGrid science gateways and their impact on science. Computer, 41(11):32{41, 2008.

[12] David R. Maidment (Ed.): CUAHSI Hydrologic Information System: Overview of Version 1.1, The University of Texas at Austin , July 12, 2008

[13] High Level Expert Group on Scientific Data: Riding the wave - How Europe can gain from the rising tide of scientific data; Final report of the High level Expert Group on Scientific Data, October 2010

[14] Franzoni, Chiara and Sauermann, Henry: Crowd Science: The Organization of Scientific Research in Open Collaborative Projects (April 8, 2013). Research Policy, http://dx.doi.org/10.2139/ssrn.2167538

[15] PARADE: The 2009 White Paper "Strategy for a European Data Infrastructure" of the European data initiative PARADE (Partnership for Accessing Data in Europe), 2009

[16] Felix Bachmann, Ian Foster, Andrew Grimshaw, Dave Lifka, Morris Riedel, Steve Tuecke: XSEDE Architecture, Level 1 and 2 Decomposition, February 2012

[17] WIRADA: Achievements, 2013; http://www.csiro.au/~/media/CSIROau/Flagships/Water%20for%20a%20Healthy%20Country%20Flagship/WIRADA/WIRADA_Achievements_Brochure.pdf

[18] Horsburgh, Jeffery S., Tarboton, David G., Maidment, David R., Zaslavsky, Ilya: A relational model for environmental and water resources data; Water Resources Research 44(5), 1944-7973; 2008

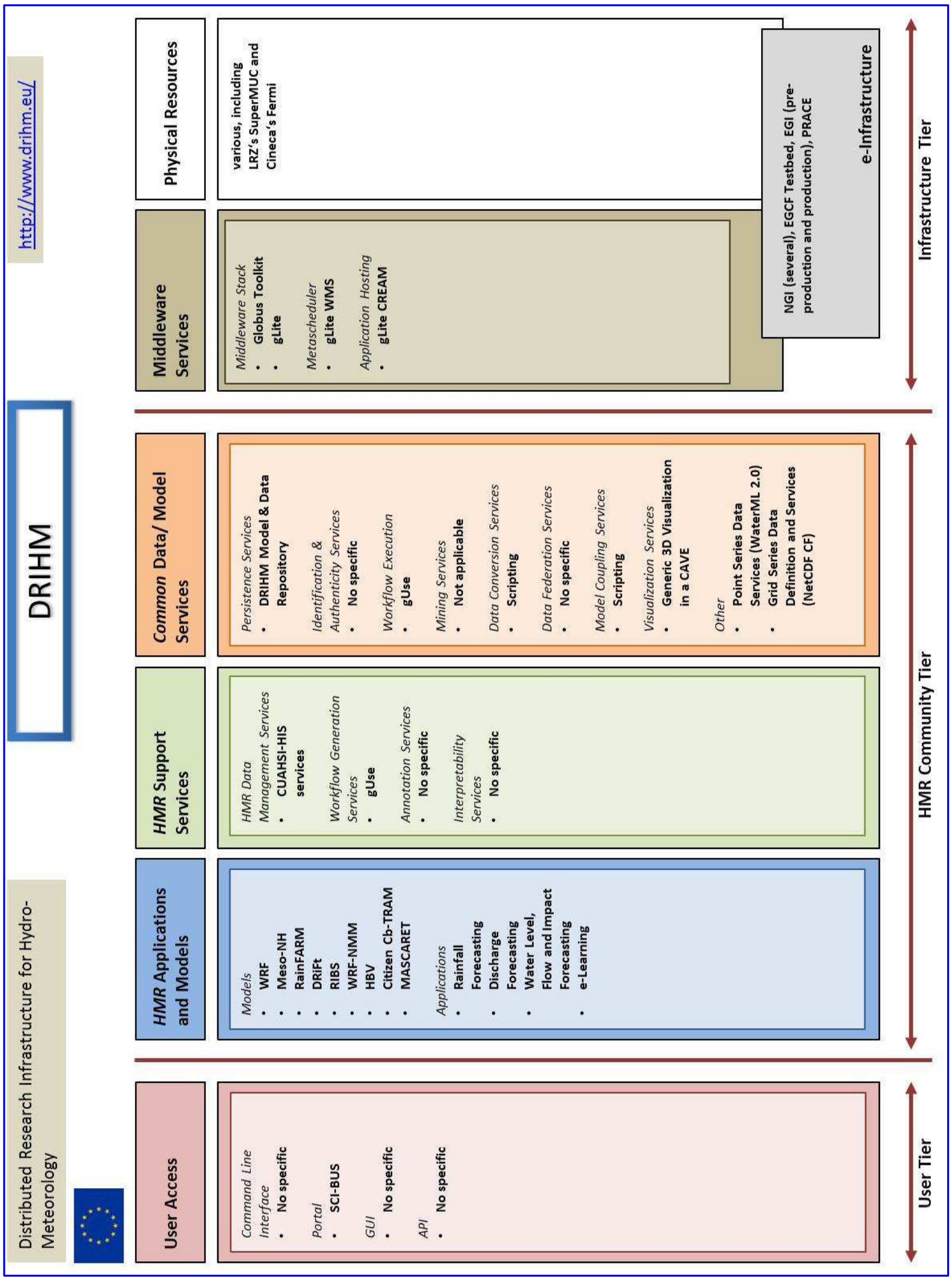
[19] Joris Borgdorff, Jean-Luc Falcone, Eric Lorenz, Carles Bona-Casas, Bastien Chopard, Alfons G. Hoekstra: Foundations of distributed multiscale computing: Formalization, specification, and analysis; J. Parallel Distrib. Comput. 73 (2013) 465–483

[20] Susan M Cuddy and Peter Fitch: Hydrologists Workbench – a hydrological domain workflow toolkit; In Proceedings of the 2010 Fifth Biennial Meeting of the International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Ottawa, Canada, 2010

Appendix

This appendix contains the specific mappings of selected e-Infrastructures onto the generic framework of Figure 2 in both a graphical and a more descriptive tabular form.

Appendix 1. DRIHM



Category	Assessment
User Access	DRIHM does not provide any specific user clients. Rather, the default clients as delivered with the various tools and models will be used. However, DRIHM provides a portal interface to the scientific workflows. The portal is based on the SCI-BUS[14] science gateway customisation methodology which itself is built on top of the generic gUSE/WS-PGRADE portal family.
HMR Applications and Models	DRIHM aims at "operating" several HMR workflows as described in detail in [1]. The context of these flows is given in Figure 4 (courtesy Q. Harpham). Implementing the workflows requires several HMR models to be coupled: • Large scale meteorological models like WRF, WRF-NMM, Meso-NH • Small scale meteorological models like RainFarm • Hydrologic models like HBV, DRiFt, RIBS • Hydraulic models like MASCARET, Delft3D-Flow • Impact models • Nowcasting models like Cb-TRAM These models are coupled into workflows rainfall forecasting, discharge forecasting, and water level, flow and impact forecasting. In addition, models and model couplings are used for training and education purposes.
HMR Support Services	DRIHM does not provide any specific annotation and interpretability services. Rather, the services as delivered with the tools will be used. Workflows can be generated with gUSE, a generic graphical user interface to create and run workflows on various infrastructures. For HMR specific data management purposes DRIHM leverages the

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

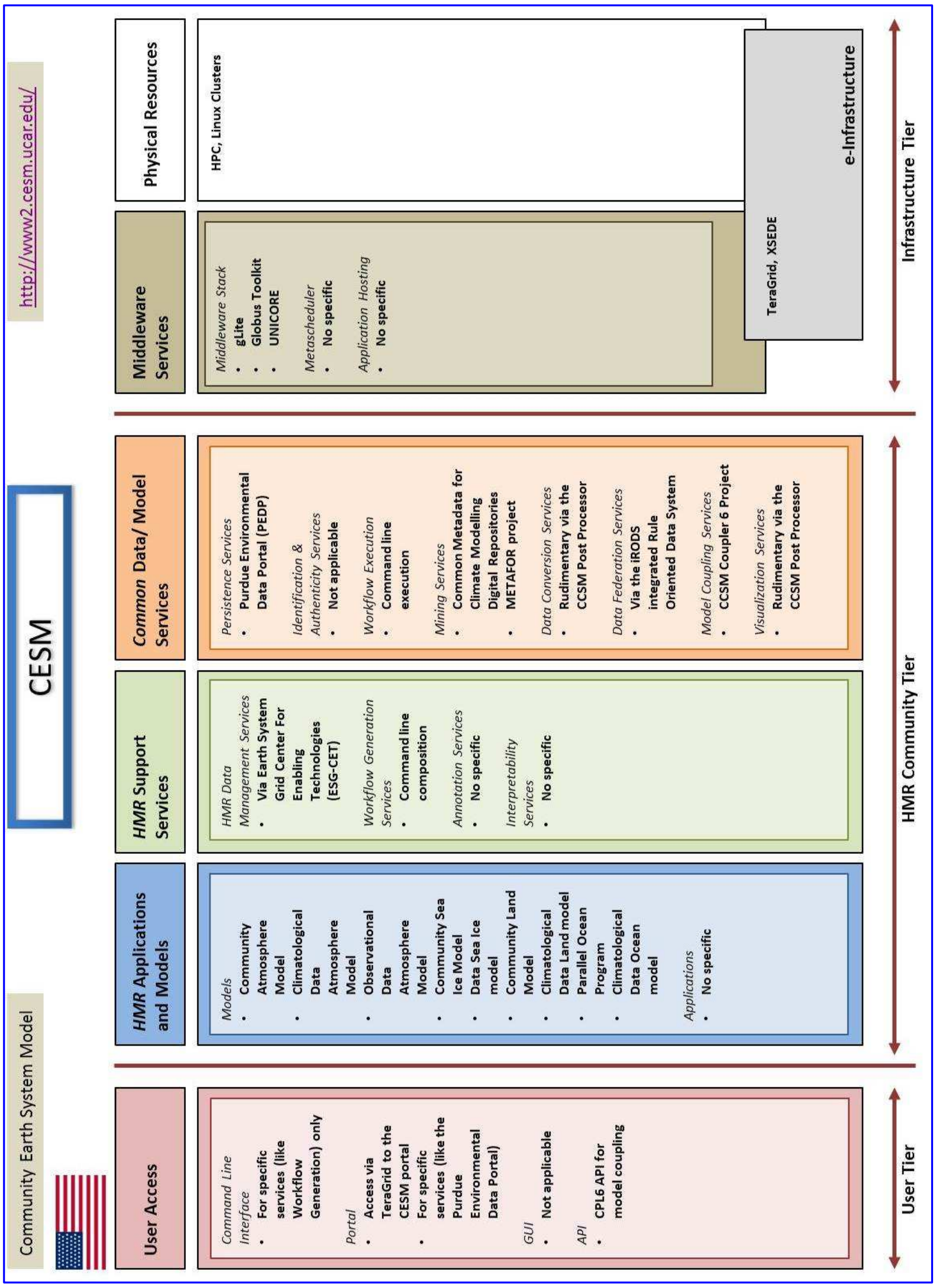
Category	Assessment
	services provided by CUAHSI-HIS (see section 3.2.4).
Common Data/Model Services	DRIHM uses a dedicated and straight forward model and data repository until an EGI provided repository is available (see right hand side of Figure 5). While the workflow execution is again based on SCI-BUS and gUSE, data conversion and model coupling is as well scripted as the 3D visualization services in the LRZ CAVE (https://www.lrz.de/services/v2c_en/). Note that DRIHM supports both the WaterML 2.0[15] standard of the Open Geospatial Consortium (OGC) for encoding of hydrological and hydrogeological observation data and the description of time series, and the NetCDF (Network Common Data Form) set of libraries for grid series data and their Climate and Forecast (CF) conventions[16].
Middleware Services	As a Grid-based project DRIHM leverages the middleware services, metaschedulers and application hosting environments as packaged with the Globus Toolkit and gLite (see left hand side of Figure 5.)
Physical Resources	DRIHM deploys into national Grid initiatives (e.g., the Italian Grid, the Polish Grid, or the German Grid), the European Grid Infrastructure (EGI), and the PRACE partnership. In addition, for basic testing purposes, DRIHM uses the test bed provided by the European Globus Community Forum (EGCF)[17].

15 http://www.opengeospatial.org/projects/groups/waterml2.0swg

16 http://www.unidata.ucar.edu/software/netcdf/

17 http://www.egcf.eu/

Appendix 2. CESM



Category	Assessment
User Access	User access to the various CESM models with relevance for HMR is provided by the TeraGrid CESM portal or via the CPL6 API supporting model couplings.[18] Currently, the primary distribution systems for CESM data are the Earth System Grid Center for Enabling Technologies (ESG-CET)[19] and the Graphical Information System (GIS) portal of the University Corporation for Atmospheric Research (UCAR).
HMR Applications and Models	CESM allows applications use Community Atmosphere Models, Climatological Data Atmosphere Models, Observational Data Atmosphere Models, Community Sea Ice Models, Data Sea Ice Models, Community Land Models, Climatological Data Land Models, Parallel Ocean Program, and Climatological Data Ocean Models.
HMR Support Services	The CESM workflows are scripted model couplings, which are executed on specific machines. During the course of a CESM run, the model components integrate forward in time, periodically stopping to exchange information with a coupler component. The coupler receives fields from the component models, computes, maps, and merges this information, then sends the fields back to the component models. The coupler brokers this sequence of communication interchanges and manages the overall time progression of the coupled system. CESM does not provide any specific annotation, interpretability, and authenticity services.
Common Data/Model	The CESM "eco system" provides various functionalities to share, mine,

[18] http://www.cesm.ucar.edu/models/ccsm3.0/cpl6/cpl6_api.pdf

[19] http://esg-pcmdi.llnl.gov/

Category	Assessment
Services	federate data and to couple models: • The Purdue Environmental Data Portal (PEDP) project provides a gateway to access various datasets managed by the Purdue TeraGrid multidisciplinary data framework. It allows researchers to perform map based data search, data browsing, metadata display, metadata query, data download and data visualization. CESM represents just one data set (see https://gridsphere.rcac.purdue.edu:8453/gridsphere/gridsphere). • The Metafor project[20] describes climate models and associated data in the standardized Common Information Model (CIM) which allows for simple data mining. • The iRODS project supports building sharable collections from data distributed across file systems and tape archives using the iCAT Metadata Catalog.[21] In addition, CESM uses NetCDF as the standard data format for all output data. Thus, all CESM components will create NetCDF output history data and all post-processed CESM data will be made available in NetCDF format. Additionally, all CESM NetCDF output history data will comply, to the extent possible, with the Climate and Forecast (CF) metadata convention[22].

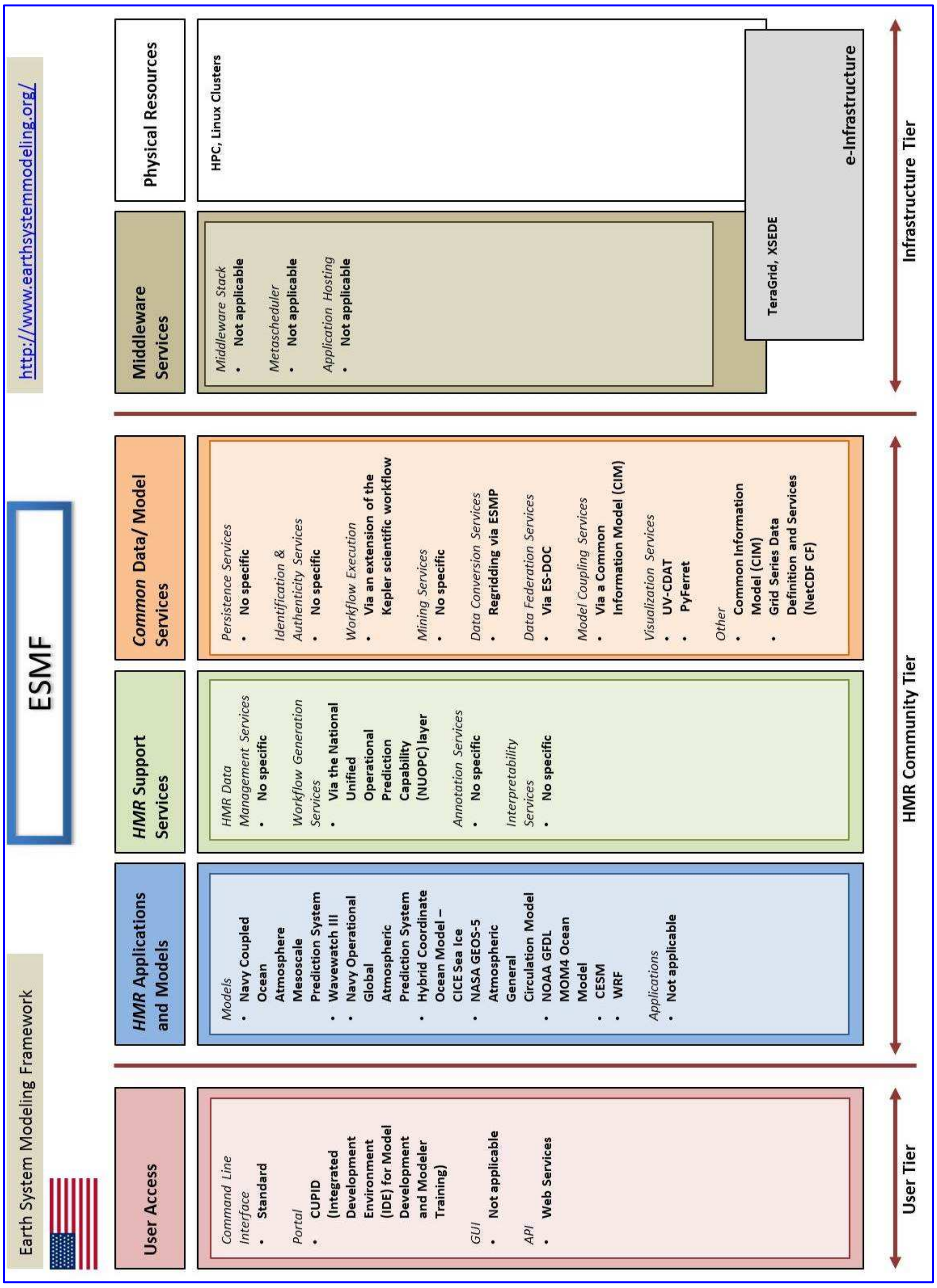
[20] http://metafortrac.badc.rl.ac.uk/trac

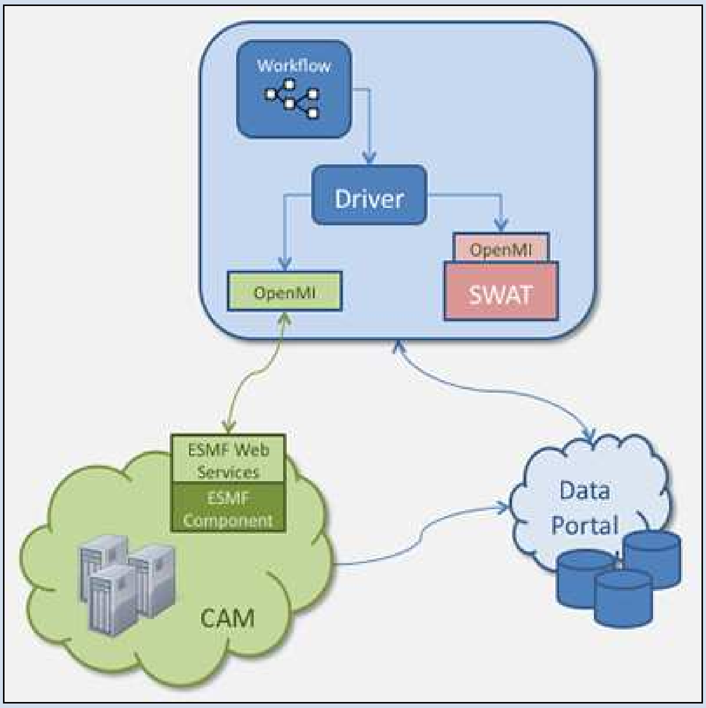
[21] https://www.irods.org/pubs/iRODS_Fact_Sheet-0907c.pdf

[22] http://cf-pcmdi.llnl.gov/

Category	Assessment
Middleware Services	Access to these resources is via Web Services CESM is providing the interfaces for. The service interfaces support the operations like case creation, case configuration, case submission, status tracking, run time information query, and data access. For more information we refer to https://code.google.com/p/cesm-ws/.
Physical Resources	CESM uses resources provided by the TeraGrid/XSEDE infrastructure.

Appendix 3. ESMF



Category	Assessment
User Access	Access to the Earth System Modeling Framework is typically done using command line clients. However, the Cupid project[23] is creating a software development and user training environment for climate models by providing an Integrated Development Environment (IDE) based on the Eclipse framework ESMF also offers a Web Service interface for model coupling via OpenMI as shown in the following figure[24]  **Figure 11: ESMF System Architecture**

[23] http://www.earthsystemcog.org/projects/cupid/

[24] Source: http://www.earthsystemcog.org/projects/esmfwebservices/architecture

Category	Assessment
HMR Applications and Models	ESMF "contains" several models of the global weather and climate modelling community (including WRF and CESM) to support coastal, hydrological, and space weather applications.
HMR Support Services	There are no specific annotation and interpretability services and no specific HMR data management services. However, for workflow generation and model coupling the National Unified Operational Prediction Capability (NUOPC) consortium is aiming at common model architecture to simplify building modelling systems. To this end, a NUOPC Layer has been defined for using ESMF. The NUOPC Layer implements a component based software architecture with four basic building blocks (drivers, models, mediators, connectors) which can be arranged in many different ways to couple models and thus to generate workflows (in their broadest sense).

Category	Assessment
Common Data/Model Services	Apart from pure model couplings, more complex workflows can be executed on the TeraGrid infrastructure by running a modified version of CCSM (Community Climate System Model) although they can be extended to use other earth system models. ESMF provides simple data conversion services (mainly regridding) via the ESMF Python interface (ESMP)[25]. Other services of the ESMF eco-system provide • Data federation via the ES-DOC[26] initiative to develop metadata services for a set of climate modelling and related projects based on the Common Information Model (CIM) that describes Earth system models • Model coupling is done either via the NUOPC Layer or via CIM or via the Model Coupling Toolkit (MCT)[27] • Visualization services are available as part of the UV-CDAT Ultrascale Visualization Climate Data Analysis Tools[28] or via the interactive visualization and analysis environment (Ferret)[29]

[25] http://www.earthsystemcog.org/projects/esmp

[26] http://www.earthsystemcog.org/projects/es-doc-models

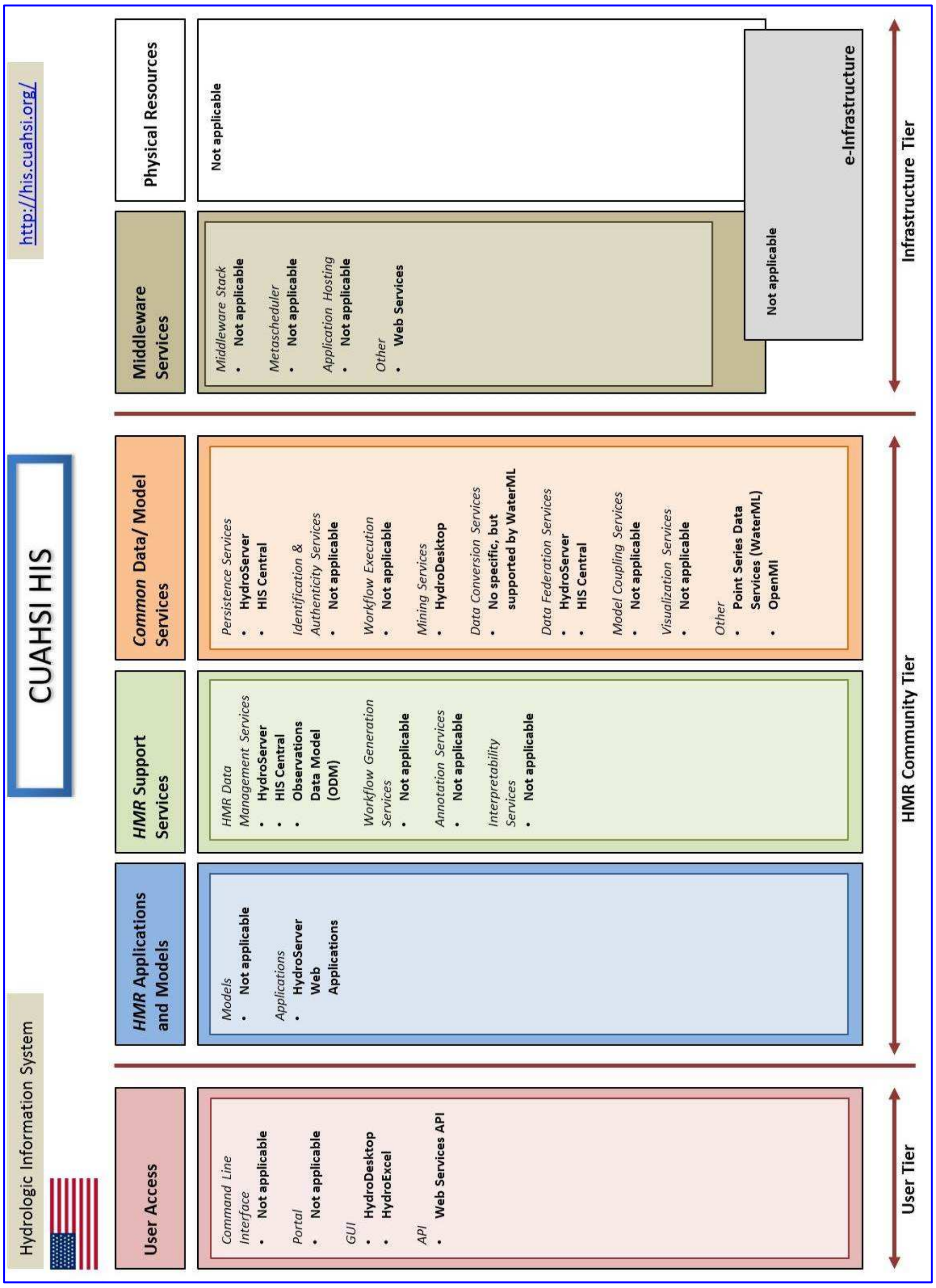
[27] http://www.mcs.anl.gov/research/projects/mct/

[28] http://uv-cdat.llnl.gov/

[29] http://ferret.pmel.noaa.gov/Ferret/home

Category	Assessment
	ESMF supports a standardized information model (CIM) the standard Grid Series Data Definition and Services (NetCDF CF).
Middleware Services	ESMF provides a so called superstructure for assembling geophysical components into applications and an infrastructure that modellers use to share common utility code for functions such as grid remapping and time management. As such, ESMF is not depending on a specific middleware, or metaschedulers or application hosting environment.
Physical Resources	ESMF may be deployed on various HPC systems and clusters, either standalone or provided by TeraGrid and/or XSEDE.

Appendix 4. CUAHSI HIS



Category	Assessment
User Access	The CUAHSI Hydrologic Information System (HIS) provides graphical user interfaces via HydroDesktop[30] clients and HydroExcel[31] clients. In addition, data access is provided via Web Services. For example, the WaterOneFlow32 Web Service defines a standard set of query functions for accessing water data, and returns the data in WaterML format.
HMR Applications and Models	HIS does not favour any specific models. Rather, it provides Web Services or Web Services APIs, tools, standards and procedures that aim at enhancing the access to data for hydrologic analysis. Typical applications thus use standard or RESTful Web Services interfaces for transferring hydrologic data between hydrologic data servers (databases) and user computers. CUAHSI-HIS uses a family of Web Services, called WaterOneFlow, that have been developed as a standard mechanism for extracting data from a data source, transforming it into a usable format and loading it in to an analysis environment. Web Services format the data as XML and the specific variety of XML that is generated by the WaterOneFlow Web Services is known as CUAHSI WaterML. Storage and publishing of hydrologic data is provided by so called HydroServers.
HMR Support Services	HydroServers are the part of the HIS system that stores and publishes hydrologic data. It is one of the three main pieces of the HIS system and it communicates with the other parts using WaterOneFlow Web

[30] http://his.cuahsi.org/hydrodesktop.html

[31] http://his.cuahsi.org/hydroexcel.html

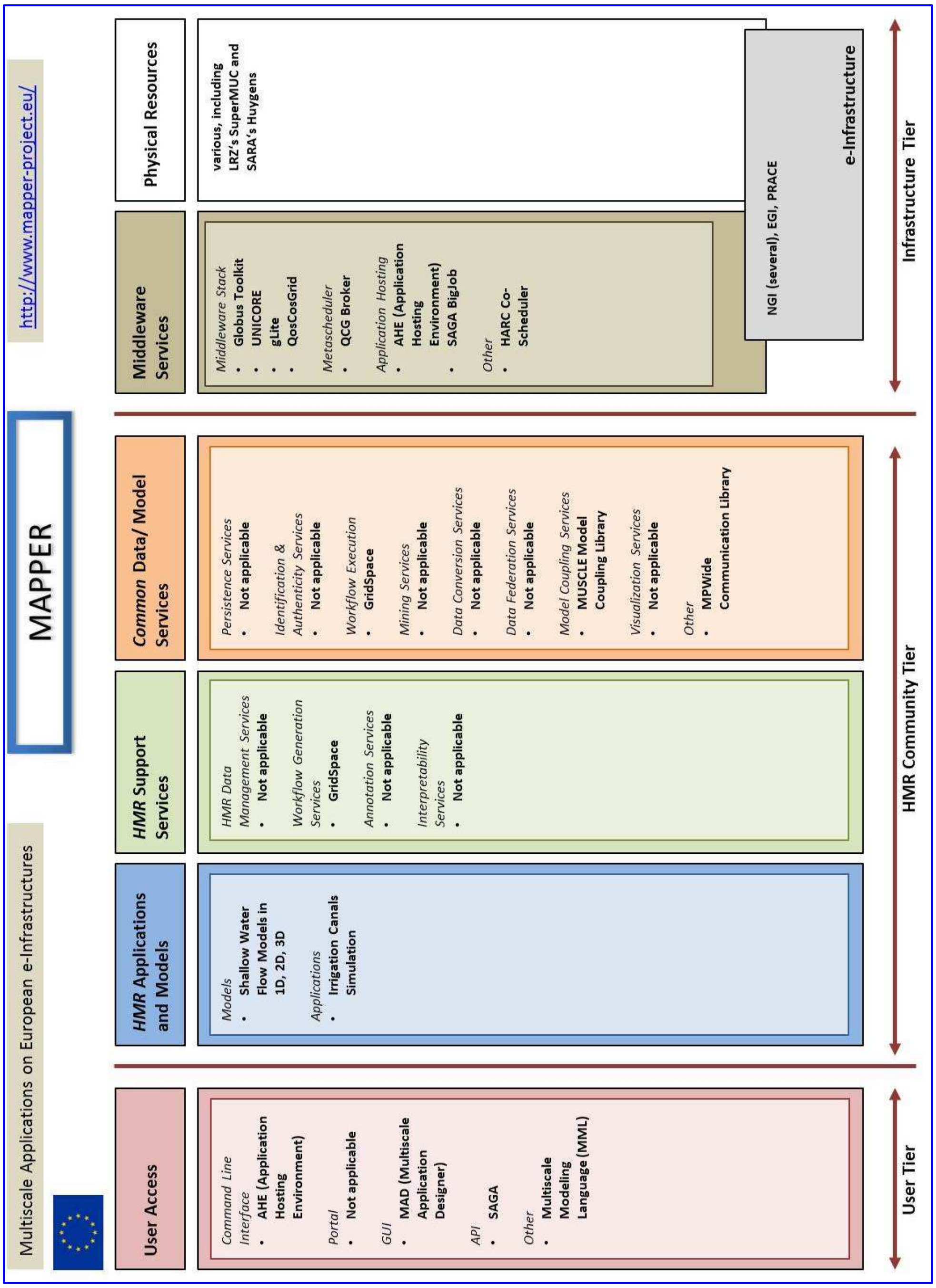
[32] http://his.cuahsi.org/wofws.html

Category	Assessment
	Services for implementing rudimentary workflows. For example, the user has some site locations and variables of interest. He sends a request to a WaterOneFlow Web Service for the data using standard functions. WaterOneFlow finds the data repository, extracts information from it, transforms the data into WaterML, and sends the result back to the user. HIS Central is the part of the system that keeps a listing of all registered Web Services, similar to the standard Web Services UDDI registry. HIS Central also maintains a metadata catalogue to enable fast data discovery and other data management services. The metadata catalogue stores the sites and a description of the data series available at those sites. A data series is all the measurements of a certain variable at a given site. HIS Central records the site and variable information as well as the begin date, end date, and number of measurements for each data series. Discovery of data is supported by a hydrologic ontology (HIS Ontology[33]).
Common Data/Model Services	Conceptually, hydrologic observations data are described by a Point Observations Information Model which itself is described using structured WaterML documents. Observations data are stored in relational Observations Data Bases which are based on the Observations Data Model (ODM) template [18]. Using ODM and the other HIS components more sophisticated HMR data management services may be implemented for data mining and data conversion.

[33] http://his.cuahsi.org/ontologyfiles.html

Category	Assessment
Middleware Services	Not applicable
Physical Resources	Not applicable, see however the installation prerequisites on http://his.cuahsi.org/ and respective subpages.

Appendix 5. MAPPER

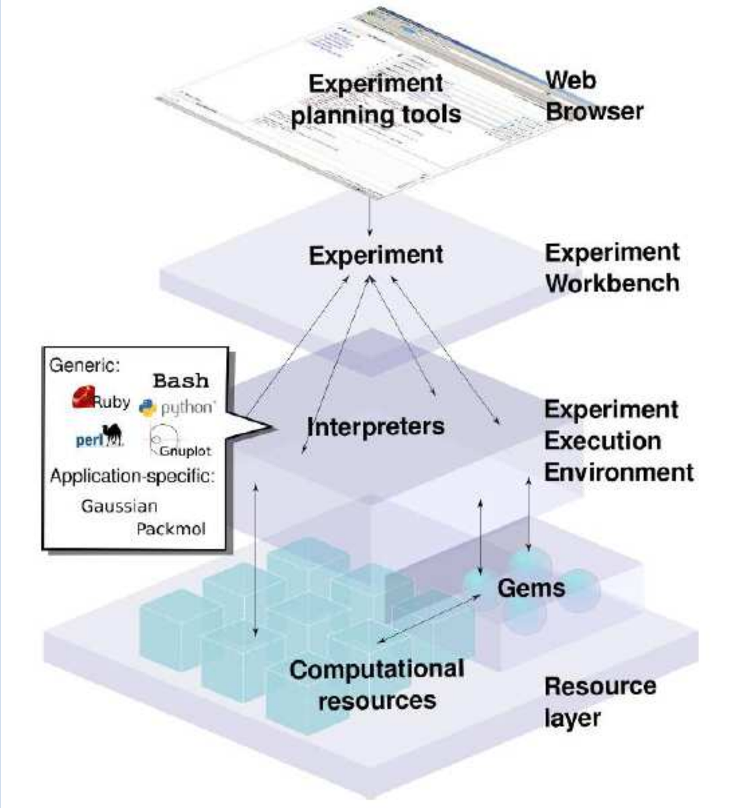


Category	Assessment
User Access	MAPPER provides no specific user interface. Rather, the project leverages product specific access to models, services and resources. These are especially: • The Application Hosting Environment (AHE) provides simple desktop and command line interfaces to run models and applications on resources provided by national and international Grids, in addition to local departmental and institutional clusters, while hiding from the user the details of the underlying middleware in use by the Grid. The AHE is able to run applications on both UNICORE and Globus Grids, meaning that a user can use a single AHE installation to access resources on NGIs, EGI and PRACE. • SAGA (Simple API for Grid Applications) is a set of simple, and generic application programming interfaces (APIs) for accessing Grid services from generic application codes, portals, and data managements systems. It is a standard developed by the Open Grid Forum (OGF) as a basis for a widely available and supported interface usable by application programmers enabling Grid functionality in their applications. • The Multiscale Application Developer (MAD) is a web application providing convenient and user-friendly set of tools allowing users to graphically compose MAPPER applications and export them to executable experiments inside the GridSpace Experiment Engine. • MAPPER provides a Multiscale Modelling Language (MML) for specifying the coupling architecture of multiscale simulations. For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.

Category	Assessment
HMR Applications and Models	MAPPER provides generic services and tools for various scientific disciplines. Of relevance for this report is the hydrologic irrigation canal application. The goal of this application is to develop hydrodynamic submodels in order to simulate a network of canals or any water course that need to be controlled and managed to produce irrigation water, electricity, etc. Due to the size of an irrigation network and the large variation in the flow complexity across different sections, some parts of the canal section can be described with 1D shallow water based models, whereas other sections need a 3D, free-surface hydrodynamic model (FS3D) to properly capture the flow properties. For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.

Category	Assessment
HMR Support Services	MAPPER provides with GridSpace 2[34] an Experiment Workbench designed to suit the requirements of domain scientists and allowing them to exploit distributed computing platforms, including Grids. GridSpace 2 is based on the notion of exploratory programming where each experiment can be decomposed into a number of *snippets*. Each snippet may be written in a different programming language. Moreover, the Workbench enables users to execute entire workflows or just selected snippets. The web portal of the Experiment Workbench assists users in iteratively developing simulation workflows with the use of popular scripting languages. The Experiment Execution Environment, which is a layer underneath the Experiment Workbench, evaluates snippets and executes them on remote sites when needed. The general architecture is shown in Figure 12. For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.

[34] http://dice.cyfronet.pl/gridspace/

Category	Assessment
	 **Figure 12: GridSpace 2 Architecture**

Category	Assessment
Common Data/Model Services	While MAPPER does not offer any particular services regarding data management, it does so for model management and workflow management. The latter is covered by GridSpace 2 (see above). The former is available by the Multiscale Coupling Library and Environment (MUSCLE). MUSCLE[35] provides the means to couple distinct computational models of multiple disciplines. Its aim is to provide a uniform platform to implement submodels using varying programming languages, execute them on heterogeneous machines and couple them across diverse networks. In addition, it considers the temporal and spatial scale at which the submodels are operating. MUSCLE allows users to specify at execution time which submodel should run on which machine. Finally, MAPPER uses MPWide[36] as a communication library intended for message-passing between supercomputers with the ability to explicitly define the communication method and configuration for individual network paths, and combine these customized configurations with an MPI-style API. As a result, MPWide allows simulations to run efficiently across heterogeneously configured wide area networks. For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.
Middleware	MAPPER tools are independent of any middleware technology although

[35] http://muscle.berlios.de/

[36] http://castle.strw.leidenuniv.nl/software/mpwide.html

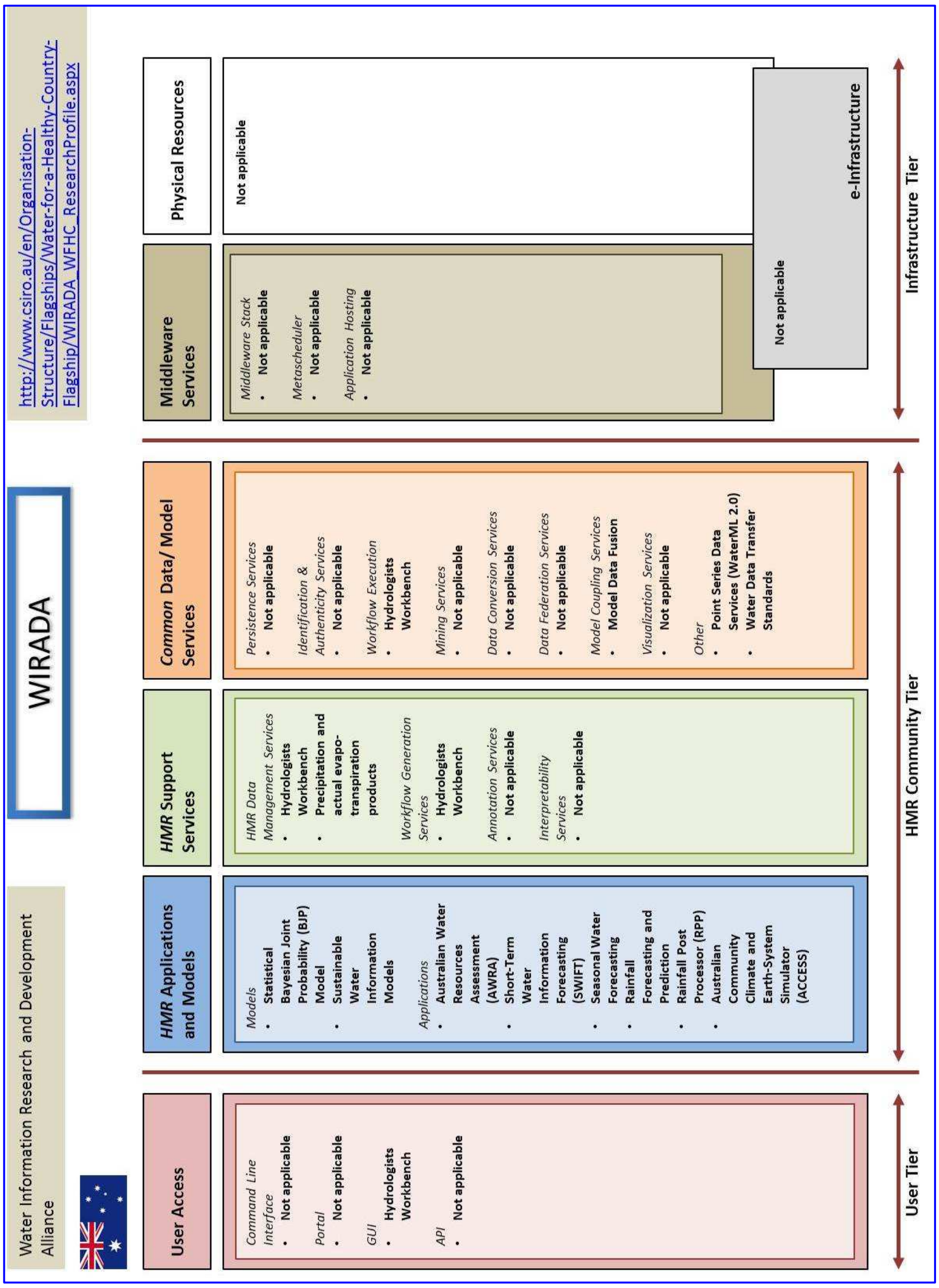
Category	Assessment
Services	it favours the QosCosGrid middleware approach[37]. QosCosGrid was designed to build multi-layered infrastructure being capable of dealing with computationally intensive large scale simulations. The developed middleware enables computing clusters in different administrative domains to be virtually welded into a single powerful computing resource that can be treated as a quasi-opportunistic supercomputer the computational power of which exceeds the power offered by a single administrative domain (data centre). The QosCosGrid infrastructure was primarily designed and developed in the QosCosGrid FP 6 project and then adopted and extended in PL-Grid national Grid initiative. MAPPER also promotes HARC, the Highly Available Robust Co-scheduler[38], developed at Louisiana State University, which empowers users by allowing them to make reservations of time on compute, storage and network resources, guaranteeing when their simulation will run. This ability of co-reservation is essential when running coupled models across different computational resources. For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.
Physical Resources	MAPPER services and tools have been deployed on various HPC systems (e.g., LRZ's SuperMUC, SARA's Huygens) in the NGI/EGI context and PRACE (see also here https://wiki.egi.eu/wiki/MAPPER-PRACE-EGI_Task_Force_%28MTF%29-II).

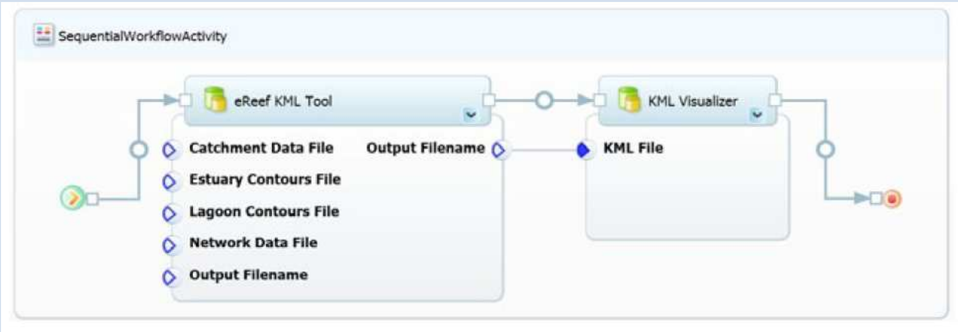
[37] http://www.qoscosgrid.org/trac/qcg

[38] http://www.realitygrid.org/middleware.shtml

Category	Assessment
	For more information we refer to the MAPPER deliverables at http://www.mapper-project.eu/web/guest/documents.

Appendix 6. WIRADA



Category	Assessment
User Access	The primary access point to the WIRADA system is the Hydrologists Workbench[39] which facilitates the automation of common workflow processes to access, calibrate and use hydrological models and data. The Workbench rolls out an integrated hydrological modelling desktop application that provides firstly the ability to discover and visualise spatio-temporal data from the Australian Water Resources Information System (AWRIS) and other sources; secondly, it provides interfaces to external services and tools that manipulate and visualise hydrological data; finally, it provides interfaces to hydrological models to integrate and orchestrate modelling tasks as part of larger workflows.  **Figure 13: Workbench Example[40]**
HMR Applications and Models	Based on a Statistical Bayesian Joint Probability (BJP) Model and Sustainable Water Information Models (SWIM)[41], several HMR relevant

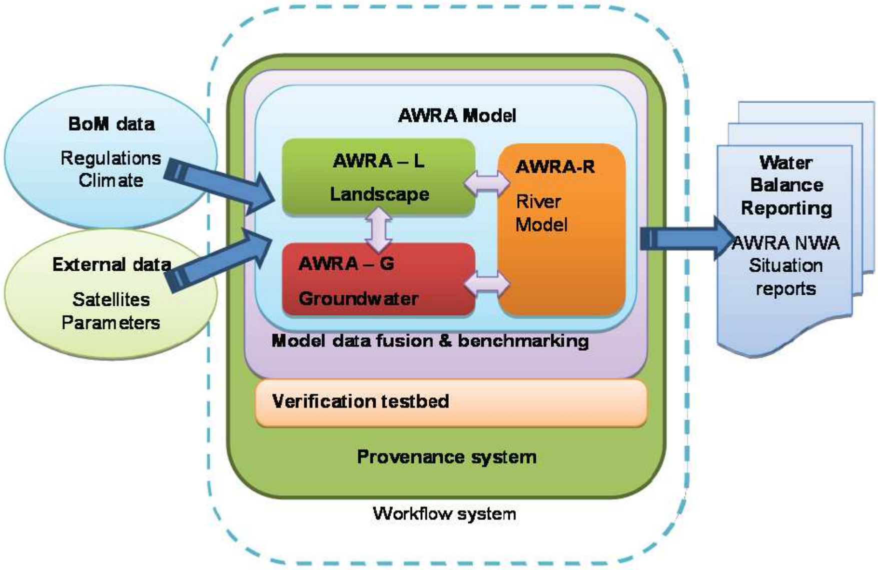
39 http://www.csiro.au/Organisation-Structure/Flagships/Water-for-a-Healthy-Country-Flagship/Integrated-Water-Resources-Management/Hydrologists-workbench_CLW_Project.aspx

40 Source: http://www.bom.gov.au/water/about/waterResearch/document/project1.2.pdf

Category	Assessment
	applications are supported. The main applications are: • Australian Water Resources Assessment (AWRA) • Short-Term Water Information Forecasting (SWIFT) • Seasonal Water Forecasting • Rainfall Forecasting and Prediction • Rainfall Post Processor (RPP) • Australian Community Climate and Earth-System Simulator (ACCESS)
HMR Support Services	Besides the Workbench for data management and workflow generation there are no specific WIRADA HMR support services available. An exception is the Gridded Foundation Data Products for developing systems to produce national data sets of daily precipitation, evapotranspiration and flood extents[42].
Common Data/Model Services	While the Hydrologists Workbench supports the execution of workflows, the workflows themselves are based on model-data fusion techniques. Model–data fusion aim at reconciling model and observed data in ways that respect their inherent uncertainties, and ensure the best available information is used to estimate water balance. Consequently, this involves identifying key data sets and determining how they link to the model; comprehensively considering model and observational errors across space and time; and choosing model–data

41 http://www.clw.csiro.au/publications/waterforahealthycountry/wirada/factsheets/wfhc-WIRADA-factsheet-SWIM.pdf

42 http://www.clw.csiro.au/publications/waterforahealthycountry/wirada/factsheets/wfhc-WIRADA-factsheet-precipitation-evapotranspiration.pdf

Category	Assessment
	fusion methods that are fit for purpose and computationally feasible. Figure 14[43] sets the context.  **Figure 14: Australian Water Resources Assessment Modelling System** WIRADA supports the WaterML 2.0 standard and the Water Data Transfer Format (WDTF)[44].
Middleware Services	Not applicable as the services do not require Grid contexts
Physical	No specific, installation requirements may be found in the respective

[43] Source: http://www.bom.gov.au/water/about/waterResearch/document/project3.1.pdf

[44] http://www.bom.gov.au/water/standards/wdtf/index.shtml

Category	Assessment
Resources	service descriptions